

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 11, 2003, 11:55:44 ; Search time 39 seconds
(without alignments) 3739.250 Million cell updates/sec

Title: US-10-009-782-1
Perfect score: 3299
Sequence: 1 gaattccactgacgcgcga.....ccctgactacgagaagctt 1758

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delpop 6.0 , Delpext 7.0	

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_n2p.model -DEV=yip
-O/C992.1/USPPO.spool/US10009782/runat_07052003_122517_23123/app_query.fasta.1.1927
-DB=SwissProt_40 -QMT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNIT=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10009782.ecgn.1.1.19.0runat_07052003_122517_23123 -MCP=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEDUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-NARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELDP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2506	76.0	483	1	NDAD_ALCXX
2	1363.5	41.3	488	1	NDAD_ALCXX
3	1000.5	30.3	488	1	NDAD_ALCXX
4	321.5	9.7	660	1	YHLL_EBV
5	312	9.6	660	1	YHLL_EBV
6	296.5	9.0	1367	1	AMTH_YEAST
7	289.5	8.9	611	1	YTI3_MYCTU
8	284.5	8.6	5179	1	MOC2_HUMAN
9	284.5	8.6	1733	1	VNDA_PRRVA
10	283.5	8.6	437	1	YGY3_HALSO
11	274	8.4	1733	1	VNDA_PRRVA
12	270	8.2	1464	1	CA11_HUMAN
13	269.5	8.2	1027	1	CAFE_RIFPA
14	268.5	8.1	1460	1	CA11_CANRA
15	267	8.1	825	1	SE5_RAT
16	264.5	8.0	1466	1	CA13_HUMAN
17	264	8.0	1464	1	CA13_MOUSE
18	263	8.0	1453	1	CA11_MOUSE

19	261	7.9	779	1	CA11_BOVIN	P02453 bos taurus
20	259.5	7.9	1453	1	CA11_CHICK	P02457 gallus gall
21	259	7.9	1255	1	MOC1_HUMAN	P15941 h. mucin 1 p
22	252.5	7.7	1459	1	CA12_MOUSE	P28481 mus musculu
23	244.5	7.4	1418	1	CA12_HUMAN	P02458 homo sapien
24	244	7.4	1461	1	IE18_PRRVE	P1675 pseudorabie
25	243.5	7.4	1372	1	CA21_MOUSE	O01149 mus musculu
26	242	7.3	555	1	GP1_CHLRE	O94P6 chlamydomon
27	240.5	7.3	1262	1	CA13_CHICK	P12105 gallus gall
28	239.5	7.3	1355	1	CA21_RANCA	O42350 rana catesb
29	239	7.2	1446	1	IE18_PRRVA	P33479 pseudorabie
30	239	7.2	1806	1	CA1B_HUMAN	P12107 homo sapien
31	238.5	7.4	825	1	SE5_RAT	O63003 rattus norv
32	238.5	7.2	1049	1	CA13_BOVIN	P04258 bos taurus
33	238.5	7.2	2944	1	CA17_HUMAN	O02388 homo sapien
34	237.5	7.2	671	1	CA11_RAT	P02454 rattus norv
35	237	7.2	620	1	EXTR_TORAC	P13983 nicotiana t
36	235.5	7.1	1650	1	CA2B_MOUSE	O64739 mus musculu
37	234.5	7.1	1323	1	NME4_MOUSE	O03391 mus musculu
38	234.5	7.1	1364	1	CA21_BOVIN	P02465 bos taurus
39	233.5	7.1	1372	1	CA21_RAT	P02466 rattus norv
40	233	7.1	1804	1	CA1B_MOUSE	O61245 mus musculu
41	232.5	7.0	1362	1	CA21_CHICK	P02467 gallus gall
42	231.5	7.0	1362	1	CA21_MOUSE	P27884 oryctolagus
43	231	7.0	2424	1	CCAA_RABIT	O9176 delonococcus
44	231	7.0	3530	1	MY15_HUMAN	O9176 delonococcus
45	229	6.9	1366	1	CA21_CANFA	O46392 canis famli

ALIGNMENTS

RESULT 1
ID NDAD_ALCXX STANDARD; PRT; 483 AA.
AC P72349: 008051:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE D-aminocyclase (EC 3.5.1.81) (N-acyl-D-amino-acid deacylase).
GN DAN.
OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Achromobacter;
OX NCBI_Taxid=515;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=96100942; PubMed=8541651;
RA Wakayama M., Katsuno Y., Hayashi S., Miyamoto Y., Sakai K.,
RA Moriguchi M.;
RT Cloning and sequencing of a gene encoding D-aminocyclase from
RT Alcaligenes xylosoxydans subsp. xylosoxydans A-6 and expression of
RT the gene in Escherichia coli.
RL Biosci. Biotechnol. Biochem. 59:2115-2119(1995).
CC - FUNCTION: HAS A WIDE SPECIFICITY; HYDROLYSES N-ACYL DERIVATIVE OF
CC NEUTRAL D-AMINO ACIDS.
CC - CATALYTIC ACTIVITY: N-acyl-D-amino acid + H(2)O -> an acid + D-
CC amino acid.
CC - COFACTOR: ZINC.
CC - SUBCELLULAR LOCATION: Cytoplasmic.
CC - SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
EMBL: S80683; AA83581.1; -
EMBL: D45918; BA808349.1; -
KW Hydrolase; Zinc.

OY	937	GGAGCCACCATATACACTCGGAGCAAGCCTTCCCAGAACTGCAGGCCGCGACCTGGAT	996
Db	301	GlyAlaThrIleIleIleThrTProSylsTroPheProGluLeuSerGlyAhaAspLeuAsp	320
OY	997	GAAGTCGGCGCCGAGCGCGGCAAAATCCAACTACACGTGTGGCCGAGCTGCAGCGGCC	1095
Db	321	GluValAlaIlaIlaGluArgGlyLysSerLysTrpAspValAlaProGluLeuGlnProIla	340
OY	1057	GGCGGCACATCTATTATATATGAGAGAACACCAGCTGCAGGCGATCTTGCGTTGGCCG	1111
Db	341	GlyAlaIleTyTPheMetMetAspLurProAspValAlaInArgIleLeuAlaPheGlyPro	360
OY	1117	ACCATGATCGGCTCCGAGCGGCTCCGCGCACACAGCGGCCGCGATCCGCGCCTGTGGGG	117
Db	361	ThreMetIleGlySerAspGlyLeuProHisAspLunArgProHisProAlaGlyeIutPrGly	380
OY	1177	ACCTTCGCGCGGGTGTCTGGGCACTATAGCGCGCAGCTGGGCGCTGTTCCTCCGTGAGAG	123
Db	381	ThrPheProAlaGlyValLeuGlyLysIstYrAlaTrgAspLeuGlyLeuPheProLeuGluThr	400
OY	1237	GGCGTAGTAGATGATACCGGCGCTACCGCGCGCGCGGCGATTCGCGCTGGCGCGGCGAG	129
Db	401	AlaValTrpLysMetThrGlyLeuThrAlaAlaArgPheIleuAlaGlyAhaGlyGln	420
OY	1297	CTGCAGCGCGGGTACTTCGCCGCGACCTGGTGTTCGACCCGCGCACGCTGGCGATACC	135
Db	421	LeuGlnAlaGlyTYrPheAlaAspLeuValAlaPheAspProAlaThrValAlaAspThr	440
OY	1357	GCCACCTTCGAACACCCCTACCGAGCGCGCGCGCGGCGATTCGATCCGTACGTCACCGG	141
Db	441	AlaThrPheGluIuhSProThrGluArgAlaIlaIleIleHisSerValIlyrValaAngly	460
OY	1417	GGCGCGCTGTGGCAAGAGAGCGCTTCACCGCGCGACGATCCGCGCTGTCTGCAGCC	147
Db	461	AlaProValTrpGingIngInaIlaPheThrGlyGlnHisAlaGlyAlrValaIleuAlarG	480
OY	1477	ACGGCGCGCC 1485	
Db	481	ThrIlaIla 483	
RESULT 2			
NDDD	ALCXH		
ID	NDDD	ALCXH	STANDARD; PRT; 498 AA.
AC	P94212;		
DT	15-JUL-1998	(Rel. 36, Created)	
DT	15-JUL-1998	(Rel. 36, Last sequence update)	
DT	15-JUL-1998	(Rel. 36, Last annotation update)	
DE	N-acyl-D-aspartate deacylase (EC 3.5.1.83) (N-acyl-D-aspartate amidoD-hydroxide).		
OS	Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxidans).		
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;		
OX	Achromobacter.		
OX	NCBI_TaxID=515;		
EN	[1]		
RN	SEQUENCE FROM N.A.		
RC	SPRAIN-A-6;		
RA	Makayama M., Matnabe E., Takenaka Y., Miyamoto Y., Tau Y., Sakai K.,		
RA	Moriyuchi M.,		
RT	"Cloning, expression, and nucleotide sequence of the N-acyl-D-		
RT	aspartate amidoD-hydroxide gene from Alcaligenes xylosoxydans subsp.		
RT	xylosoxydans A-6";		
RL	J. Ferment. Bioeng., 80:311-317,(1995).		
RP	[2]		
RP	CHARACTERIZATION.		
RC	SPRAIN-A-6;		
RX	MEDLINE=93372486; PubMed=7763985;		
RA	Moriyuchi M., Sakai K., Matsuno Y., Maki T., Makayama M.;		
RT	"Purification and characterization of novel N-acyl-D-aspartate		
RT	amidoD-hydroxide from Alcaligenes xylosoxydans subsp. xylosoxydans		
RL	A-6";		
RL	Biosci. Biotechnol. Biochem. 57:1145-1148(1993).		
CC	-I-CATALYTIC ACTIVITY: N-acyl-D-aspartate + H(2)O -> carboxylate + D-		
	aspartate.		

Db 235 I L e V a l p r o A l a m e t G l u G l u A l a L e u L e u I l e G l y A r g G l u L e u A s p C y s A r g V a l 254

-1- CATALYTIC ACTIVITY: N-acyl-D-glutamate + glutamate.

	RESULT	3			
	NDED_ALCXX				
ID	NDED_ALCXX	STANDARD;	PRT;	488	AA.
AC	p94211;				
DT	15-JUL-1998 (Rel. 36,	Created)			
DT	15-JUL-1998 (Rel. 36,	Last sequence update)			
DT	15-JUL-1998 (Rel. 36,	Last annotation update)			
DE	N-acyl-D-glutamate deacylase (EC 3.5.1.82) (N-acyl-D-glutamate amidohydrolyase).				
OS	Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxidans).				
OC	Bacteria; Proteobacteria; Beta subdivision; Alcaligenaceae;				
CC	Achromobacter.				
OX	NCBI_Taxid=515;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN-A-6:				
RX	MEDLINE=96015170; PubMed=8537313; Wakayama M., Ashika T., Miyamoto Y., Yoshikawa T., Sonoda Y., Sakai K., Moriguchi M.;				
RT	Primary structure of N-acyl-D-glutamate amidohydrolyase from				
RL	Alcaligenes xylosoxydans subsp. xylosoxydans A-6.";				
CC	J. Biochem. 118:204-209(1995).				
CC	- CATALYTIC ACTIVITY: N-acyl-D-glutamate + H(2)O = carboxylate + D- glutamate.				

CC -1- COFACTOR: ZINC.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC EMBL: D50061; BAA08778.1;
 DR InterPro: IPR002195; Dihydrocoetase.
 DR Pfam: PF00744; Dihydrocoetase; 1.
 KW Hydroxylase; Zinc.
 SQ SEQUENCE 488 AA; 51492 MW; F5A1B8315BEA167F CRC64;
 Alignment Scores:
 Pred. No.: 1,93e-35 Length: 488
 Score: 1000.50 Matches: 223
 Percent Similarity: 59.38% Conservative: 62
 Best Local Similarity: 46.46% Mismatches: 190
 Query Match: 30.33% Indels: 5
 DB: 1 Gaps: 4
 US-10-009-782-1 (1-1758) x NDED_ALCXK (1-488)
 QY 52 CAGCCCTTCGACCTGCTGCTCGCGGGGACCCCTCATGAGCGAGCAACACCCCGGG 111
 DB 3 GILVSLAASPLEUVALLEGLUNGLYTPVAILLEASPGLYLEUNGILYPROARG 22
 QY 112 CGCGCGCGCGACCTGGCGCTGCGCGCGACCCATCGCCCGCATGCGGCGATCTGCGAC 171
 DB 23 AGAGALAAASPVAILGLYLEARGILYGLUARGLLEALALEALILEGLYASPLEUERALA 42
 QY 172 GCGCGCGCGACACCGCGGCTGCTGCGCGCTGCTGCGCGCGCGCGCGCTTCAATGAC 231
 DB 43 ALAPROALLASPARARGLEUASPARAAGLYARGILEVALAIPROGLYPHEILEASP 62
 QY 232 TCGCACACCCAGACGACACTGCTGCTGCGCGCTGCGCGACGACCGCGCAAGATGCG 291
 DB 63 THRLISGLYHLSASPARLEUMETHEVALIGLULYSPROGLYLEUNGILTRIPLYSHTSER 82
 QY 292 CAGGCGCTGACACGCTGCTGCGCGCAATGCGCGCATGCGCGCGCGCGCTG 348
 DB 83 GINGLYLLETHTSERVALVALVALGLYASNCYSGLYILESERGLYALAPROALPROLEU 102
 QY 349 CAGCGCAACCGCGCGCGCGCTGCTGCGCGCAAGCGGCTTACCGTTTCGAG 408
 DB 103 PROGLYASHTHRLAALALEUALALEUEN-----GLYASPSERPROLEUPHEALA 120
 QY 409 CGCTTCGCGACTGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCTG 468
 DB 121 ASPMETALAMETLYRPHGLYALALEUNLUALAGLNAIPROMETILEASVVALAALA 140
 QY 469 ATGGGCGCGCATTCACGCTGCGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCG 528
 DB 141 LEUVALIGLYHLSASPARLEUARGLEUALALEALAMETARPARPROALALEGLNPROSER 160
 QY 529 GAGCGAGAAATCG 588
 DB 161 ALALYSGILUNARGALAMETGLUNARGLEUALAASPARALALUGLUALAGLYALAVAL 180
 QY 589 GGCATTTGCGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648
 DB 181 GLYPHESERTHRLYLEUALATYGLNPROGLYLYVALAAGLUGLUNLALAGLUNLEUASP 200
 QY 649 GAGGTGCG 708
 DB 201 GLYLEUALAARGVALAALALEALARGLYALALEUHTHTSERHLSILEARGASGLU 220
 QY 709 GCGGAGCACATCGTGGCGCGCGCTGGAGAAACCTTCGCGCATGCGCGCGCGCG 768

DB 221 GLYASPARALVALIGLUALALEVALASPGLYLEUALLEUALGLYARGARGHTHRLYCY 240
 QY 769 CCGGTGCTATCTTCGACCAACAGTGCATGCGCGCGCGCGCGCGCGCGCGCG 828
 DB 241 ALATHVALLEUSERHLSHLSYSCYSMETETPROALASHTPGLYASERIALA 260
 QY 829 ACGTGCCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885
 DB 261 THRLAALASHTLEASPARVALARGALAAAGLYVALASPVALLALEUASPLETYR 280
 QY 886 CCTACGTGCGCGCGCTCCACATGCTCAAGCAGACCGCGCTGCTGCGCGCGCG 945
 DB 281 PROTYRPROGLYSEHTSERHTHRLLEULEIEPROGLUNARGALASPGILILEASPLE 300
 QY 946 ATCATCCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1005
 DB 301 ARGILEHTHTPSEHTHTHRLPROHISPROGLUCYGLYGLNLSERLEUALLEULLEA 320
 QY 1006 GCGGAGCGCGCGCAATTCAGACAGACGCGTGGCGCGCGCGCGCGCGCGCG 1065
 DB 321 ALAARGTPGLYCYASPARALVALTHRLAALARGARGLEUCYASPROALGLYALILE 340
 QY 1066 TACTTCATGATGAGCAACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1125
 DB 341 TYRPHLEALAMETASPGILUNASGLUNLVALARGARGHTHRLHLSGLUCYSCYMETVAL 360
 QY 1126 GCGTCCGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1185
 DB 361 GLYSERASPGLYLEUPROASHTPARHLSHLSHLSHLSHLSHLSHLSHLSHLS 380
 QY 1186 CGGTGCTGCGCGCGCAATTCAGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1245
 DB 381 ARGVALLEUGLYARGTYVALARGLUNLGLULLEUHTHRLLEUGLUALALEVALA 400
 QY 1246 AAGATGACCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1305
 DB 401 LYSMETHTHRLALEUAPROALARGVALAPHEGLYLEUALASPARAGLYARGLEUALAVAL 420
 QY 1306 GCGTACTGCG 1365
 DB 421 GLYALATTPALASPVALLVALVALPHEASPARHLSHLSHLSHLSHLSHLSHLS 440
 QY 1366 GAACACCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1425
 DB 441 ASPALAPROHTHRLLEUALASERIALGLYILELUNHLSVALLEUVALASNGLYCYAALVAL 460
 QY 1426 TGCGAAGACGAGCGGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1485
 DB 461 PHE---PROGLNALAPROPROSERHLSARGPROGLYARGILEUALARGALASPARSER 479
 RESULT 4
 YHLL_EBV
 ID YHLL_EBV STANDARD; PRT; 660 AA.
 AC P03181;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Hypothetical BHFL1 protein.
 OS Epstein-Barr virus (strain 95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seglin C.,
 RT Tuftnell P.S., Barrett B.G.,
 RT "DNA sequence and expression of the 95-8 Epstein-Barr virus genome.";
 RC Nature 310:207-211(1984).
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch.)

CC EMBL: V01555; NOT_ANNOTATED_CDS.
 DR PIR: A03742; Q0B83.
 KM Hypothetical protein; Early protein; Repeat.
 FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
 FT REPEAT 149 273 1.
 FT REPEAT 274 398 2.
 FT REPEAT 399 523 3.
 FT REPEAT 524 648 4.
 SQ SEQUENCE 660 AA: 66244 MW: 86DA1D67A37152A2 CRC64;

Alignment Scores:

Pred. No.: 4,436-07 Length: 660
 Score: 321.50 Matches: 187
 Percent Similarity: 31.83% Conservative: 18
 Best Local Similarity: 29.04% Mismatches: 256
 Query Match: 9.75% Indels: 183
 DB: 1 Gaps: 29

US-10-009-782-1 (1-1758) x YH1_EBV (1-660)

QY 39 CCAATCCGATCCGACCCCTGACCTGCT----- 68
 DB 113 ProGlnArgSerProArgThrArgGlnAlaGlyTyrAlaLeuGlyGlnGlySerAlaGly 132
 QY 69 -----GCTGGCGGGGGGCGGACCCCTCATCGAGCGGACACACCCCGGGGGGGCGCGCA 122
 DB 133 LeuGlySerAlaGlyProArgProAlaPheGlnAlaGlnTrpSerAlaGlnAsn 152
 QY 123 CTTGGCGGTGGCGGGGCGGACCGCATCGCGCATCTGCGAGCCCGCGCGCA 182
 DB 153 ProGlyCysPro-----ArgThrTrpArgArgArgSerGlyAlaGlnArgGly 168
 QY 183 CACCGGGGTGGACGTGTGGGCTGTGGTGGCGCGCGCTCATGACTGCACACCA 242
 DB 169 HisPro----- 170
 QY 243 CGAGCACAACCTACCTGTAGCGGTGGCGCATGAGCCCAAGATCTCGAGGGCGTAC 302
 DB 171 -----ProProGlyAlaGlyGlnArgProSerGlyProThrGlyAlaG 185
 QY 303 CACGGTGTACACGGGCAATGCGGCATCAAGCGCGCGCGTGGCGCACGCCAACCGCGC 362
 DB 186 -----ProAlaAlaProGlyAlaProGlyThrPro 195
 QY 363 CGCCCGCTGGACCTGTGTGAGCAGAGCGGCTGTACCGTTTCGAGCGCTTCGCGCACTA 422
 DB 196 AlaAlaProGlyProGlyGly--GlyAlaAlaValPro--SerGlyAlaThrProHisP 214
 QY 423 CCTGGACGGCTTGGGGCA-----CGCC 446
 DB 214 roGlnArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlnArg 234
 QY 447 GGGCGCGCTCAACGCGCGCTGTATGGTGGCCCATTCACCGCTGCGCGCGCGGTATGCG 506
 DB 234 InGlnProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysPro----- 249
 QY 507 GGAATTGACGCGGCGCGCACCGACGAGGAATCGCGGCGCATCGGGAGCTGGCCGAGGA 566
 DB 250 -----AlaGlyProProProThrArgSerGly----- 258
 QY 567 AGCGATGGCGAGGGGCGCATCGGCATTTGACCGCGGCTTCTACCCCGCGCGCGCG 626
 DB 259 -----AlaAlaAlaGlnArgThrHisArg-----ArgProProGlyC 271
 QY 627 CGGCACACCGAGAGATCATCGAGGTGTGCGCGCGCTGAGCGCGCATGCGCGCATCTA 686
 DB 627 ----- 686

DB 271 yspPro-----ArgSerAlaArg-AsnProGlyCysProArgThrTrpArg----- 285
 QY 687 CCGCACCCATGCGCGAGAGGCGGACATCTGCGCGC-----GCTGGAGA 737
 DB 286 -----ArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGly 300
 QY 738 AACCTTCGCATCGCGCGCGAGCTGAGCTGCCGTGTGATCTGCACACAGATCAT 797
 DB 301 GlnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaPro----- 317
 QY 798 GGGCGACGCCATTTGCGCGCGCTGCGGAGACGCTGGCGGATGAGCGCGCGCATGCGC 857
 DB 318 -----GlyThrProAla 321
 QY 858 GCGCGAGAGCTGTCTGAGCGCGATTCCTACGTGGCGCGCTCCACCATCTCAAGCA 917
 DB 322 AlaProGlyProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGlnArg 341
 QY 918 GGA-----CGCGTGTCTGTGGCGGAGCGGACGACCATCATCTGCTGCA 962
 DB 342 GlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlnArgGlnArg 361
 QY 963 GCGCTTCGCGGAGTGGCGGCGGACCTGGATGAACTCGCGCGGAGCGCGCAATC 1022
 DB 362 ArgLeuProGlnAspLeuAlaAla-----GlnArgCysProAlaGlyProPro 378
 QY 1023 CAAGTACAGCTGTGTCGCGGACCTGACGCGCGCGCGCATCTTACTGATGAGAGCA 1082
 DB 379 ProThrArgSerGly-----AlaAlaAlaGlnArgThrHisArgArgProProGlyCys 396
 QY 1083 ACCCGACGTGACGCGCATCTGCGCGTGGCGCGGACCATGATCGGCTCGAGCGGCTGC 1142
 DB 397 ProArgSerAlaArgAsnProGlyCys--ProArgThrTrpArgArgArgSerGlyAla 415
 QY 1143 GCAGCAGAGCGCC-----GCATCCGCGCTGTGGCGACCTT 1181
 DB 416 GlnArgGlyHisProProProGlnArgGlnArgGlnArgProSerGlyProThrGlyAlaArg 435
 QY 1182 CCGCGGGTGTGGGCA-----CTATCGCGCGCATCTGCGGCT- 1220
 DB 436 ProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAla 455
 QY 1221 GTTCCCGCTGGA-----GACGGGCGTATGAGAGTACCGCGCTGAC 1262
 DB 456 ValProSerGlyAlaThrProHisProGlnArgGlySerGlyProAlaAspProProAla 475
 QY 1263 CGCGCGCGCTTTCGCGCGGCGCGCGCG-----GCACCTGCA 1301
 DB 476 AlaAlaArgLeuProProGlnArgGlnArgProArgLeuProGlnAspLeuAlaAla 495
 QY 1302 GCGCGGTACTTTCGCGGACCTGTGTGTCGACCGCGGCGGCGGCGGATACCGGCAC 1361
 DB 496 GlnArgCysProAlaGlyProProProProThrArgSerGlyAlaAlaGlnArgThrHis 515
 QY 1362 CTTCGAAACCCCTACGAGCGGCG--CGCGGCAATCCATTCGCGTACGTCACAG--G 1415
 DB 516 ArgArgProProGlyCysProArgSerAlaArgAsnPro--GlyCysProArgThrTrp 535
 QY 1416 CCGCGGTCTGCGCAAGACAGCGGTTCACCGCGGACGATGCGCGCGCGGTCTGCAAG 1475
 DB 535 rgnArgArgSerGlyAlaGlnArgGlyHisPro----- 545
 QY 1476 CACGCGCGCTGAGCGCGCGCGCGCGCTTACATTCGCGGCGGAGGCGCGGCTGCGC 1535
 DB 546 -----ProProGlyAlaGlnArgGlnArgProSerGlyProThr--GlyGlyArgProAla 563
 QY 1536 CCGCGTCCCAACCTGAGCAAGCAACCGCTACATGCGCC-----CTCCCTCCGC 1583
 DB 563 IAProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProS 583
 QY 1584 TCGCAATFACGCGCGCGCATCTGCGGCAAGAGTATGAGGCGCGCGCTGCGCGC 1643
 DB 583 erGlyAlaThrProHisPro--GlnArgGlySerGlyProAlaAspProProAlaAlaAla 602

```

QY 1644 CGAGCCGAGAGCCGCGGAAATGACCCCTGCAGACCT-----GTGCGAGCCGAGCG 1694
DB 603 ArgLeuProProGluArgGlnGlnIupProArgLeuProGlnAspLeuAlaAlaGlnArg 622
QY 1695 ---CATGCGGCTGTCGAGCCCTGCAGAGCCGAGCTGGCGCCAGATGCGCTGAGTACGA 1751
DB 63 CysProAlaGlyProProProThrProThrSerGlyAlaAlaAlaGlnArgThrHisArgArg 642

RESULT 5
YHL1_EBV STANDARD: PRT: 660 AA.
ID YHL1_EBV
AC P03181
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLF1 protein.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
CX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Bartell B.G.,
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."
RL Nature 310:207-211(1984).
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: V01555; NOT_ANNOTATED_CDS.
DR PIR: A03742; Q0BE3.
KM Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D6A7152A2 CRC64;

Alignment Scores:
Pred. No.: 1,1e-06 Length: 660
Score: 312.00 Matches: 181
Percent Similarity: 31.59% Conservative: 30
Best Local Similarity: 27.10% Mismatches: 232
Query Match: 9.62% Indels: 225
DB: 1 Gaps: 31
US-10-009-782-1 (1-1758) x YHL1_EBV (1-660)
QY 1753 TCTGTAGCTCAGAGGCGATCTGCCAGCT-----CGGCGT 1718
DB 87 SerArgGlnSerArgThrGlnGlnAlaGlnAlaAspHisAlaHisSerAspPro 106
QY 1717 TGGACAGGTCGAGACCCGATGCCGCTGGCTGCAGACAGCTTTCGAGGTCATTTCC 1658
DB 107 ThrGlyGlyCysSerAspProGlnArgSerProArgThrArgGlnAlaGlyTyrAlaLeu 126
QY 1657 GGGCCCTTGCCTCGGCGGCGAGCGGCGCCATCACTTCCGCCACAGATTCGGGTG 1598
DB 127 GlyGlnGlySerAlaGlyLeuGlySerArgGlyProArgProHisProAlaPheGlnVal 146
QY 1597 -----GGGCGGTATGC----- 1586
DB 147 GlnTrpSerAlaArgAsnProGlyCysProArgThrTrpArgArgSerGlyAlaGln 166

```

```

QY 1585 -----GAGCGGAGGAGGAGGCGCATGACGCT----- 1559
DB 167 ArgGlnHisProProProGlyAlaGlyGlnArgProSerGlyProThrGlyArgPro 186
QY 1558 -----TTGCGTACAGGCTGGAGGAGGCG----- 1535
DB 187 AlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaVal 206
QY 1534 -----GCGACGCGCGCGCTTCCAGCGGATGTAGAGGCTGGCGCGCGCTCAGCGG 1481
DB 207 ProSerGlyAlaThrProHisProGlnArg-----GlySerGlyProAlaAsp 222
QY 1480 CCGTGGCGGCGAGCAGCG-----GCGCGG 1457
DB 223 ProProAlaAlaAlaValGluLeuProProGlnArgGlnIupProArgLeuProGlnAspLeu 242
QY 1456 CATGCTGGCGCGTAAACGCTCTCTTCCAGACCGCG----- 1418
DB 243 AlaAlaAlaGlnArgCysProAlaGlyProProProThrArgSerGlyAlaAlaGln 262
QY 1417 CGCGCTTGACGTACAGGAAATGATCCGCGCGCGCTCGTAGGT-----GTTGCA 1364
DB 263 ArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysArgArg 282
QY 1363 AGTGGCGGATTCG-----CCACCGTGGCGCGGTGCAACA 1328
DB 283 ThrTrpArgArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArg 302
QY 1327 CCACGAGTCGCGGAGTACCGCGCTGCACCT-----GCCGCGCGCGG 1283
DB 303 ProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAla 322
QY 1282 CCAGCGCGAGAGCGCGCGCGCTGACGCGCGATTCATCCGCGCTCCAGCGGA 1223
DB 323 ProGlyProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGlyArgGly 342
QY 1222 ACAGCGCCAGGTGCGCGCATAGTGCCTCA----- 1193
DB 343 SerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlnArgGlnIupProArg 362
QY 1192 -----GCACCGCGGAGAGTGCCTCCAGCGCGGATGCGGCGCTCGT 1148
DB 363 LeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArgSer 382
QY 1147 CGTGGCGAGCGCGCTGCGAGCCGATCATGCTGGCGCGGAGCGCGAGATGCTGCGACGT 1088
DB 383 GlyAlaAlaAlaGlnArgThrHisArg-----ArgProProGlyCysProArgSer 399
QY 1087 CGGGTTCGTCATCATGATGATAGT-----TGAGCGCGCGCGCTGCAGCTGCGGACGCGT 1031
DB 400 AlaArgAsnProGlyCysProAlaGlyThrTrpArgArgArgSerGlyAlaGlnArgGlyHis 419
QY 1030 CGTACTGTGATTTGGCGCGCTGCGCGCGCATTCATCCAGCTGCGCGCGCTGCGG 971
DB 420 ProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaPro 439
QY 970 GGAAGGCTTGCACACAGTATATGCTGCTCGCGCGACAGCAGCGCTCTGCT--- 914
DB 440 GlyAlaProGlyThr-----ProAlaAlaProGlyProGlyGly 452
QY 913 -----TGACATGCTGGAGCGCGCAGTGGGATAGCGGTCCAGAGAGA 869
DB 453 GlyAlaAlaValProSerGlyAlaThrProHisProGlnArgGlySerGlyProAlaAsp 472
QY 868 CGTCTCGGCGG-----CCATGGCGCGCTCGATACGG 836
DB 473 ProProAlaAlaAlaArgLeuProProGlnArgGlnIupProArgLeuProGlnAspLeu 492
QY 835 GCGAGCTGTCGCGCGAGCGCGCAATTTGGCGCTGCGCGCATGACCTTGCTGCGAGATCA 776
DB 493 AlaAlaAlaGlnArg-----CysProAlaGlyPro 502

```

```

OY 775 CCAACGGCAGCT-----CCAGCTCGGGCGGCGATGCGGAAGTTTCTCCAGCGCGGCCA 722
Db 503 ProProthrinArgSerGlyAlaAlaAlaGlnArgHisAlaGlyProProGlyCysPro 522
OY 721 CGATGTCCTGCGCTTCCTCCGCGATGTGGTGCGGTAGATCCCGCATGCGCGCTCAGCG 662
Db 523 ArgSerAlaArg-----Asn 527
OY 661 GCGGCGACACCTCATGATCTCTTCGTTGGTGGCGCGCGCGCGCGGTAGAGCGGC 602
Db 528 ProGlyCysPro-----ArgThrTrpArg-----ArgArgSerGlyAlaGlnArg 542
OY 601 CCGGCGAAATGCGGATGCGCGCGCGCGCATGCTTCCTGCGCGCGCGCGCGCATGCGCG 542
Db 543 -GlyHisProProGlyAla-----GlyAlnArgProSerGlyPr 556
OY 541 CGATTCCTGCTGCGTGGCGCGCGCGCTCAGATGCGCGCATGCGCGCGCGCGCGCTTG 482
Db 556 o-----ThrGlyGly-ArgProAlaAlaProGlyAlaProGlyThrProAlaAla 572
OY 481 AATGCGCCA-----CCATACAGCGCGCGCTTGA 455
Db 573 ProGlyProGlyGlyGlyAlaAlaAlaProSerGlyAlaThrProHisProGlyAlaGly 592
OY 454 CCGGCGCGCGCGCGCGCGCGCGCGCATGCGCGCATGCGCGCGCGCGCGCGCGCTAG 395
Db 593 SerGlyProAlaProProAla-----AlaAlaAlaArgLeuProPro 606
OY 394 AGCGCGCTTCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338
Db 607 GlnArgGlnGlnProArgLeuProGlnAlaAlaAlaGlnArgCysProAlaGly 626
OY 337 CCAAGCTGATGCGCGCATGTCGCCCTGACACCGCTGTCAGCGCGCGCGCGCGCGCTGG 278
Db 627 Pro-----ProPro-----629
OY 277 TCATGTCGCGACGCGCTGAGCAGTAGTGTGCTGCGGTGGTGGCGCATGAAGCGCG 218
Db 630 -----ThrArgSerGlyAlaAlaAlaGlnArgThrHis 640
OY 217 GCGCGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 167
Db 641 ArgArgProProGlyCysProArgSerAlaArgHisProGlyCys-----655
OY 166 ACAGATGCGCGCGCGCGCG 149
Db 656 ---ProArgThrTrpArg 660

RESULT 6
AMYL_YEAST STANDARD: PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucanase SI/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
glucosylase) (1,4-alpha-D-glucan gluconhydrolase).
GN STAI OR STAI OR MALD OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288c / AB972;
RA Barrett B.G., Badcock R., Bankier A.T., Bowman S., Brown D.,
RA Church C.M., Connor K., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gault S., Hamlyn N., Horsnell T.S., Hunt S., Jagsels K., Jones M.,
RA Louis E., Lye G., Moulé S., Moulé T., Odell C., Pearson D.,
RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE-87194600; PubMed-3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAI ";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE-89031230; PubMed-3141213;
RA Pardo J.M., Imanz E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STAI and SGA genes
RT from Saccharomyces cerevisiae";
RL FEBS Lett. 239:179-184(1988).
CC -1 CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1 SIMILARITY: TO S. POMBE SPBC215.13.
CC -1 SIMILARITY: SOME, TO S. POMBE SPCC285.13C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z38061; CA86176.1; -
DR EMBL; M16164; AAA35014.1; -
DR EMBL; M16165; AAA35015.1; -
DR EMBL; X13857; CA32069.1; -
DR PIR; B26877; B26877.
DR PIR; A26877; A26877.
DR PIR; S48478; S48478.
DR SGD; S0001458; MDC1.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 817
FT CARBOHYD 874 874
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CMC64;

Alignment Scores:
Pred. No.: 4.6e-06 Length: 1367
Score: 296.50 Matches: 139
Percent Similarity: 38.72% Conservative: 72
Best Local Similarity: 25.50% Mismatches: 266
Query Match: 8.99% Indels: 68
DB: Gaps: 13

US-10-009-782-1 (1-1758) x AMYL_YEAST (1-1367)
OY 8 ACTGATCGCGGAGAGAGATTCATGTCATGCGCATGTCACCGCTTCGACCTGC 67
Db 345 ThrSerThrThrGlnSerSerSerAlaProValProThrProSerSerThrThr 364
OY 68 TGCCTGCGGGGGGACCTATGCGAGCGACACACCGCGGGGGCGCGCGACCTGG 127
Db 365 GlnSerSerAlaProValThrSerThrThrThrThrThrThrThrThrThrThr 382
OY 128 GCGTGGCGCGGCGCGCATGCGCGCGCATGCTGTGCGAGCGCGCGCGCACACC 187
Db 383 ValThrSerSerThrThrThrThrThrThrThrThrThrThrThrThrThrThr 398
OY 188 GGGTGCAGCTGCGCGCGCTGTGCTCCCGCGCGCTTCATCGACTGCACA----- 238
Db 399 SerSerSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 418
OY 239 -----CCACGACGAGAACACTGCTCAGCGCGTGCAGCATGACGCCCAAGATCGCG 292

```

Db 419 SerAlaProValThrSerSerThrThrGluSerSerSerAla-----ProValThrSer 436
 QY 293 AGGCGTACACGACGAGTGTACAGGCGCAATTGCGCATGAGCGTGGCCGCTGGCGACG 352
 Db 437 SerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAla 456
 QY 353 CCAACCGCGCCGCCCTGAGCTGCTGACGAGGAGGCGCTCTTACCGTTGAGCGCT 412
 Db 457 ProValProThrPro-----SerSerSerThrThrGluSerSerSer 470
 QY 413 TCGCG-----ACTACGTGAGCGCGTGGCGGCGACCGCGCGCGCTGACAG 460
 Db 471 AlaProValThrSerSerThrThrGluSerSerSerAlaProValProThrProSerSer 490
 QY 461 CCGCGTGTATGTGGCCATTCACGCTGCGCGCGCGGCGACCGGACCTTCCAGCGG 520
 Db 491 SerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAla 510
 QY 521 CCGCGACCGACGAGAAATCGCGCGCATGCGGACCTGGCGGAGAACCATGCGACG 580
 Db 511 ProValProThr----- 514
 QY 581 GCGCATTCGCGCATTCGACCGGCGGCTTACCGCGCGCGCGCGCGCGCGCGACGAG 640
 Db 515 -----ProSerSerSerThrThrGluSerSerSerAlaProAlProThrProSerSer 533
 QY 641 AGATCATCGAGGTGTGCGCGCGCTGAGCGCGCATGCGCGCATTCACCGCGCGCGAC 700
 Db 534 ThrThrGlu-----SerSerSerAlaProValThrSer 544
 QY 701 GCGAGAGAGCGGACGACATCGTGGCGCGCGCTGAGAGAAACCTTCGCGATGCGCGCGAG 760
 Db 545 SerThrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSer 564
 QY 761 TGGAGTGCCTGCTGTATTCGACACGACAAAGCATGGCGCGCGCGCGCGCGCGCT 820
 Db 565 SerSerThrProValThrSerSerThrThrGluSerSerSerAlaProValProThrPro 584
 QY 821 CCGCGGACACGCTGCGCTGATCGAGCGCGCGCATGCGCGCGCGCGCGCGCGCGAG 880
 Db 585 SerSerSerThrThrGluSerSerSerAlaProValProThrProSerSerSerThrThr 604
 QY 881 CGTATCCTAGTGGCGCGCTCCACACATGCTTCAAGAGAGCGCGGTGCTGCGCGAG 940
 Db 605 GluSerSerSerAlaProAlaProThrProSerSerSerThr-----Glu 620
 QY 941 GCACCATCATCATCT-----GGTGAAGCGCTTCCCGAAGC 976
 Db 621 SerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProValProThr 640
 QY 977 TGAAGCGGCGCGACCTGATGAGTCCGCGCGCGCGCGCGCAATCCAGTACAGACCTG 1036
 Db 641 ProSerSerSerThrThrGluSerSerSerAlaProValProThrProSerSerSerThr 660
 QY 1037 TGGCCGAGTGTGACCGCGCGCGCGCGCATCTTACATGAGAGAGAGCGCGAGC 1096
 Db 661 ThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSer 680
 QY 1097 GCATTCCTGGCGTCCGCG-----CGACCATGATCGGCTCCGACGCGCTGCGCGACG 1150
 Db 681 AlaProValThrSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThr 700
 QY 1151 AGCGCGCGCATCGCGCTGTGGGCACTTCCGCGGCGGTGCGGGGCGACTAGCGCGG 1210
 Db 701 GluSer-SerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAl 720
 QY 1211 ACCTGGCGCTGTCCCGCTGAGAGCGCGGTATGAGAGTACCGCGCTGACCGCGCGG 1270
 Db 720 AlaProValProThrProSerSerSerThrThrGluSerSerSerAlaPro--ValProThr 739
 QY 1271 GGTTCGCGCTGCGCGCGCGCGCGCGCGCTGACCGCGCGGTACTCGCGCGCGGTGTGT 1330
 Db 740 ProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSer 759

QY 1331 TCGACCGCGCGCGCGCGCGCGCGCGCATTCGAGAACCCCTACCGAGCGCGCGCG 1390
 Db 760 SerSerAlaPro--ValProThrProSerSerSerThrThrGluSerSerSerAlaPro 778
 QY 1391 GCATTCCTGCGGTACGTCACGCGCGCGCGGTGTGCGAAGAGACGCGCTTCACGCGG 1450
 Db 779 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSer 798
 QY 1451 AGCATGCG 1510
 Db 799 SerSerThrThrGluSerSerSerAlaProValProThrProSerSerSerSer--Asn 817
 QY 1511 CCGCGCTGACG 1570
 Db 818 IleThrSerSerAlaProSerSerSerThrProPhaSerSerSerThrThrGluSerSer 837
 QY 1571 CCGCTTCCT 1579
 Db 838 ProValPro 840

RESULT 7
 YTL3_MYCTU STANDARD: PRT; 611 AA.
 AC Q10830;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RY2913C.
 GN RY2913C OR MT2981 OR MTCY274.45C OR MTCY338.01C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (Class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID:1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV.
 RX MEDLINE-98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garler T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jags K., Kiroh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sluson J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh.
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Feldman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL; 274024; CA98380.1;
 DR EMBL; AE007120; AA947307.1;
 DR TIGR; MT2981;

DR Tuberculin; RV2913c; ..
 KW Hypothetical protein; Hydrolase; Complete proteome.
 SQ SEQUENCE 611 AA; 67205 MW; C3AA4A13B0BC34A CRC64;

Alignment Scores:

Pred. No.:	6,28e-06	Length:	611
Score:	294.00	Matches:	153
Percent Similarity:	36.48%	Conservative:	79
Best Local Similarity:	24.06%	Mismatches:	216
Query Match:	8.91%	Indels:	190
	1	Gaps:	25

US-10-009-782-1 (1-1758) x YH13_MYCTU (1-611)

```

OY 58 TTCGACCTGCTGTCGGCGGCGGACCCCTCATGACGGCAGCAACCCCGGCGGCGC 117
DB 16 TTTASPVAllelleArspgilyLeuTPPhaspilyLeuTgAsnAlaProleuThr 35
OY 118 GCCGACCTGGCGCTGGCGCGGACCGCATCGCC-----GCCATGGCGATCTGTGGAC 171
DB 36 ArgThrLeuGlyLeuArspgilyValAlaAlaThValAlaAlaGlyAlaLeuAspGlu 55
OY 172 GCCCGCGCGCACACCCGCGTGCAGCTGTGCGGCGCTGTGCGCGCGCTTCATCGAC 231
DB 56 ThrGlyCysProGluValAlaAspAlaAlaGlyLysTrpValValProGlyPheIleasp 75
OY 232 TCGCACACCCAGCACACACACTGCTGTCAGCGCTGCGGACATGACGCCCAAGATCG 291
DB 76 ValHisThrHisIleTrpAspAlaGluValLeuLeuAspProGlyLeuArgIleuSerValArg 95
OY 292 CAGGCGCTCACACAGCTGCTGTCAGCGCAATTGGCGCATCGCTGGCGCGCGTGGCGCAC 351
DB 96 HisGlyValThrThrValLeuLeuGlyAsnCysSerLeuSer-----ThrValTyr 112
OY 352 GCCAACCCGCGCGCGCGCTGACCTGCTGAGACGAAAGC----- 390
DB 113 AlaAsnSerGluAspAlaAlaAspLeuPheSerArgValGluAlaValProArgGluPhe 132
OY 391 -----GGCTCTTAACGT-----TTCGACCGCTTCGCCGACTACCTGAGCGCG 432
DB 133 ValLeuGlyAlaLeuArspgilyLeuThrProAlaGlyIleGluAla 152
OY 433 TTCGCGGCGCACCGCGCGCGCGCTGCAACCGCGCTGTATGATGGCGCATTCACGCGCGC 492
DB 153 IleAspAlaLeuProleuGlyProAsnValSerSerLeuLeuGlyHisSerAspLeuArg 172
OY 493 GCCGCGGTATCCCGGACTTGCAGCGCGCGCC-----ACCGAGAG 534
DB 173 ThrAlaValLeu---GlyLeuAspArgAlaThrAspAspThrValArgProThrGluAla 191
OY 535 GAAATCGCGGCATGCGGAGACTGCGCGGAGAAAGCCATGGCGCGCGCATCGGCATT 594
DB 192 GluLeuAlaLeuMetAlaLysLeuLeuAspGluAlaLeuGlyMetLeuGlyMet 211
OY 595 TCGACCGGCGCTTGTACCGCGCGCGCGCGCGCACACCAAGAGAGATCAT----- 647
DB 212 Ser-----GlyMetAspAlaAlaIleAspLysLeuAspGlyAsp 224
OY 648 -----CGAGGTGTGCGCGCGCGCTGAGCGCGCATGGCGCATTCACCGCACCA 695
DB 225 ArgPheArgSerArgAlaLeuProSerThrPheAlaThrTrpArgGlyLysLeu 244
OY 696 CATGGCGCGACGAGCGAGCATGCTGCGCGCGCTGAGGAAACCTTCGCGCTGCGCGCG 755
DB 245 IleSerValLeuArg-----HisArgGlyArgIle-----LeuGlnSerAlaPro 259
OY 756 CGAGCTGACGTCGCGGTGATCTCGCACACCAAGGTCATGGCGCGCACCCCAATTTCG 815
DB 260 -AspValAspAsnProValSerAlaLeuLeuPhePheLeuAlaSerSerArgGlyLeuPheAs 279
OY 816 CCGCTGCGCGC-----GAGACGCTGCGCGCT 839
DB 279 nArgArgIysGlyValArgMetSerMetLeuValSerAlaAspAlaLysSerMetProle 299

```

```

OY 840 GATC-----GAGGC 848
DB 299 uAlaValHisValPheGlyLeuGlyThrArgValLeuAsnLysLeuLeuGlySerGlnVal 319
OY 849 CGGCATGGCGCGCGGACGCTTCGCGGACCGGTAT----- 885
DB 319 LArgPheGlnHisLeuProValProPheGlyLeuTyrSerAspGlyTyrLeuAspProVal 339
OY 886 -----CCCTACGTGGCGCGCTCCACCATGCTCAAG-----CAGGACGC 923
DB 339 LpHeGluGluPheGlyAlaGlyThrAlaAlaLeuHisLeuArgAspGlnLeuGlnArgAs 359
OY 924 CGTGTCTGTGGCGGACCGCAC----- 945
DB 359 nGluLeuLeuAlaAspArgSerTyrArgArgSerPheArgArgGluPheAspArgIleLeu 379
OY 946 -----ATCATCTACCTGTGCAAGCC 965
DB 379 sLeuGlyProSerLeuThrPheHisArgAspPheHisAspAlaValIleValGluCys---Pr 398
OY 966 CTTCGCCGAACTGACGCGCGCGCGCATGATGAATGCGCGCGCGGCGGCAATCCAA 1025
DB 398 OAspLysSerLeuIleGlyLysSerPheGlyAlaIleAlaAspGluArgGly----- 415
OY 1026 GTACGACGTGTGCGCGGCTGACCGCGCGCGCGCAT-----CTACTCATGATGG 1078
DB 416 -----LeuHisProLeuAspAlaPheLeuAspValLeuValAspAs 429
OY 1079 ACGAACCCGACGTCGACGCGATCTGTGCGCGCGCGCATGATGCTGCGCGCGCC 1138
DB 429 nGlyGluArgAsnValArgTrpThrThrIleValAlaAlaHis-----ArgPr 445
OY 1139 TCGCGCACGACGACCGCGCGCGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1195
DB 445 OAsnGlnLeuAsnLysLeuAlaAlaGluProSerValHisMetGlyPheSerAspAlaGlu 465
OY 1196 GGCATATGCGCGCGCGCGCGCGCGCTTC----- 1224
DB 465 yAla-HisLeuArgAsnMetAlaPheTyrAsnPheGlyLeuArgLeuLeuLysArgAla 485
OY 1225 -----CCGCTGAGAGCGCGGCTGGAAGATGA 1252
DB 485 rGAspAlaAspArgAlaGlyGlnProPheLeuSerIleGluArgAlaValTyrArgLeu 505
OY 1253 CCGCGCTGACCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACT 1312
DB 505 hrcIyGluLeuAlaGluTrpPheGlyIle---GlyAlaGlyThrLeuArgGlnGlyAspA 524
OY 1313 TCGCGACCTGTGTGCTTGCAGCCGCGCGCGCATCGCGCGCATCCGCGCGCGCGCGCG 1372
DB 524 rGAlaAspPheAlaValIleAspProThrHisLeuAspGluSerValAspGlyTyrHisG 544
OY 1373 CT-----ACCGAGCGCGCGCGCGCA 1393
DB 544 lGluAlaValProTyrTyrGlyGlyLeuArgArgMetValAsnArgAsnAspAlaThrV 564
OY 1394 TCCATTCGCTGTCACGAGCGCGCGCGGTGCGGAAAGCGCGCGCTTCACCGCGCGAG- 1452
DB 564 aValAlaThrGlyValGlyGlyThrValValPheArgGlyGlnPheGlyGlyGlnP 584
OY 1453 -----CATGCGCGCGCGCGCTGTCGACGACGCGCGCGCGCT 1486
DB 584 heArgAspGlyTyrGlyGlnAsnValLysSerGlyArgTyr----- 597
OY 1487 GAGCCGCGCGCGCGCGCGCTTACAATCCGCGGTGAACGCGCGCGCG 1530
DB 598 -----LeuArgAlaGlyGluLeuGlyAlaAla 606

```

RESULT 8
 MUC2_HUMAN
 ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
 AC Q02817; Q14878;

FT CONFLICT 1351 1351 H -> L (IN REF. 3).
 FT CONFLICT 1412 1412 T -> S (IN REF. 3).
 FT CONFLICT 1449 1449 L -> P (IN REF. 3).
 FT CONFLICT 1504 1504 M -> T (IN REF. 3).
 FT CONFLICT 4192 4192 G -> S (IN REF. 2).
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571EB9A5663 CRC64;

Alignment Scores:

Pred. No.: 8.01e-06 Length: 5179
 Score: 289.50 Matches: 146
 Percent Similarity: 32.92% Conservative: 41
 Best Local Similarity: 25.70% Mismatches: 164
 Query Match: 8.78% Indels: 217
 Gaps: 29

US-10-009-782-1 (1-1758) x WUC2_HUMAN (1-5179)

QY 20 AAGAGAGATTCCATGTCATCCGAT-----CCAGCCCTCG 61
 Db 1383 LysilethgvalasnycyscttrpmetasplyscysilethrThrProserProPro 1402
 QY 62 ACCTGCTCTCGCGGCGGCGACCCCTCATGAGCGGAGACACACCCGGGGCGCGCG 121
 Db 1403 ThrThrThrProserProProProThrThrThrThrLeuPro-----Pro 1418
 QY 122 ACCTGGGCGTGGCGGCGGCGCATCGCGGCATCTGCGACGCGCGCGCGC 181
 Db 1419 Thr-----ThrThrProserProProThrThrThrThrProPro 1434
 QY 182 ACACCGGCGTGCAGTGTGCGGCGCTGTGTGCGCGCGCGCTTCATGACGTGACACCC 241
 Db 1435 ThrThrThrPro-----SerProProThrThrThrThrThrThrPro 1448
 QY 242 ACAGACACACTACTGCTCTAGGCGTGCAGACATAGCGCCACAGATCTCGAGGGCGTCA 301
 Db 1449 LeuProThrThrThr-----Proser 1455
 QY 302 CCAAGGTGTGACAGGGGAAATGGCGCATCAAGCTGGCGCGCGCGACGCCACCGC 361
 Db 1456 ProProProlester-----ThrThrThrThr 1463
 QY 362 CCGCGCCCTGAGACCTGTCAGACGAGGCGGCTTACGTTTGAGCGCGCTGCGCACT 421
 Db 1464 ProProProThrThr-----ThrProserProPro 1473
 QY 422 ACCTGAGCGGCTGCGGCGGCGCGCGCGCTCAAGCGCGCTGTATGTGGCCATT 481
 Db 1474 ThrThrThrProserProProThrThrThrProserProProThrThr 1489
 QY 482 CAACGCTGCGCGCGGCGTATCGGACTGACAGCGCGCGCGCGACGAGGAAATCG 541
 Db 1490 -----ThrThrThrThrProProProThrThrThrPro 1500
 QY 542 CGGCATGCGGAGACCTGCGCGAGAGACCATGGCCAGCGCGCGCATCGGATTTCCAGCG 601
 Db 1501 SerProProMetThrThrProlesterPro-----Pro 1511
 QY 602 GGGCCTTACCGCGCGCGCGCGCGCGCGCACGAGAGATCATGAGTGTGCGCGC 661
 Db 1512 AlaserThrThrThrLeuProProThrThrThrPro----- 1523
 QY 662 CGCTGAGCGCGATGCGGCGCATCATCCACACCATGGCGGAGAGGAGACATCG 721
 Db 1524 -----SerProProThrThrThrThrThr 1532
 QY 722 TGGCGCGCTGAGAGAAACCTTCGCGATCGCGCGAGCTGAGCGCGGTGATATCT 781
 Db 1533 ProProProThrThrThrProserProProThr----- 1543
 QY 782 GCGACACAAAGTATGCGCGCGCGCGCGCGCGCGCGAGAGAGCGTGCCTGA 841
 Db 1544 -----ThrThrProlester----- 1548

QY 842 TCGAGCCCGCATGCGCGCGCGAGACGTCGTCGAGCGGATCCCT---ACGTGCGC 898
 Db 1549 -----ProPro-----ThrSerThrThrThrLeuProProThrThrPro 1562
 QY 899 GCTCCACCATGCTCAAGCAGACGCGGTGCTGCGCGCGAGCACATCATCACTGCT 958
 Db 1563 SerProProProThrThrThrThr-----ProProProThr 1575
 QY 959 GCAAGCCCTCCCGCAATGAGCGGCGCGACCTGATGAAATGCGCGCGAGCGCGCA 1018
 Db 1576 ThrThrThrProserPro-----ProThrThrThr 1584
 QY 1019 AATCCAGT-----ACGAGTGTGTGCGCGAGCTGACGAGCGCGCGCGCA 1063
 Db 1585 ThrProserProProThrThrThrThrThrThrThrProProProThrThrPro 1604
 QY 1064 TCTACTCATATGAGACAC-----CCGACGTGACCGCATCTGCGCTGCGCGCA 1117
 Db 1605 ProThrThrThrThrThrThrThrProProThrThrThrThrProserProThrThr 1624
 QY 1118 CGATGATGCTCGAGCGGCTGCGCGCGCGAGAGAGCGCGCGCATCCGCGCTGCGGCA 1177
 Db 1625 ProlesterProProThr-----ThrThr----- 1633
 QY 1178 CTTCCCGCGGCTGCTGCGGCACTATGCGCGCGACTGTGCTTCCGCTGAGAGCG 1237
 Db 1634 -----ThrLeuProProThrThr 1640
 QY 1238 CGTATGAGAGATGACGGCGCTGACCGCGCGCGCTTCCGCTGCGCGCGCGCAGC 1297
 Db 1641 -----ProserProProProThrThrThrThrThrThrPro 1651
 QY 1298 TGCAGCGCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATA--- 1354
 Db 1652 ---ProProThrThrThrProserProProThrThrThrThrProserProProlester 1670
 QY 1355 -----CGGCACCTTGAAACCCCTACGAGCGCGCG----- 1387
 Db 1671 ThrThrThrProProProThrThrThrThrProserProlesterThrThrProser 1690
 QY 1388 -----CGGACATCCATCCGCT 1405
 Db 1691 ProProThrThrThrMetThrThrThrProserProThrThrThrProser-----Prole 1709
 QY 1406 ACCTCAACGCGCGCGCGCTGCGAGAGAGCGGCTTCAACCGCGCGAGATGCGCGCGCG 1465
 Db 1710 ThrThrThrThrThrProser-----SerThrThrThrProserProProThrThr 1727
 QY 1466 TGCTGACAGCAGCGCGCGCTGAGCGCGCGCGCGCGCGCGCTTACATCCGCGTGAACGGG 1525
 Db 1728 MetThr-----ThrProserProThrThrThrProserProThrThrThrMet 1744
 QY 1526 CGGCGTGGCGCGCGCTCCCAACCTGAGCAAAACGGTACATGGCGCGCTCCGCGCTC 1585
 Db 1745 ThrThrLeuProProThrThrThrSerSerProleuThrThrThrProleuProPro--- 1763
 QY 1586 GCAATACG 1609
 Db 1764 SerThrThrProProThrThrThr 1771

RESULT 9

VNUA_PRIVKA STANDARD: PRT: 1733 AA.
 ID VNUA_PRIVKA P33485.
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Probable nuclear antigen.
 OS Pseudorabies virus (strain Kaplan) (PRV).
 OC Viruses: dsDNA viruses, no RNA stage: Herpesviridae;
 OC Alphaherpesvirinae: Varicellovirus.
 RN NCBI_TaxID=33703;
 RN [1]


```

OY 249 CAACACTGCTGAGCGGACATGACGCGGACGAGATGTCGACAGGCGGTCAACGCGT 308
DB 141 -----GlyArgHisAlaSerAspArgValGlnAspGlyAlaHisPro 154
OY 309 GGCACGGCGCATGTCGGCATGACCGTGGCGGCGGACGCGGACGCGGCGGCGGCGG 368
DB 155 ArgArgGlnArgLeuArgGlnArgProArgHisAlaGly--ArgProArgArgArgGln 173
OY 369 CCGGACGTCGTCGAGGCGGCGTTCACGCTTCGAGCGGCGTTCGCGGCGTTCGAG 428
DB 174 ProProArgArgGlyArgSerArgGlyThrHisArgArgHisLeuArgGlnAlaProArg 193
OY 429 CCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 485
DB 194 ProAlaValArgGlyProAspGlnAspGlnAlaArgGlnProArgHisArg 213
OY 486 -----GCTGCGGCGG-----GCGGTCATGCGGCGGCGGCGGCGGCGG 518
DB 214 ArgGlnArgHisProProThrAlaArgAspValLeuArgGlyGlnProGlyHisGlyAsp 233
OY 519 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 563
DB 234 GlyHisHisLeuGlnGlyArgArgGlyArgProArgProGlnGlyArgGlnAlaGlyArg 253
OY 564 GGAAGCCAT-----GGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 608
DB 254 GlyAlaHisProProGlnValAlaArgAlaArgGlyLeuAlaAlaGlyGlnAlaArgGly 273
OY 609 CTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 668
DB 274 LeuProGlnProArgProGlnGlyValArgThrValHisArgGly----- 288
OY 669 CCGCGCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 728
DB 289 -----GlyArgLeuArgGlyArgValGlyGln 297
OY 729 GCTGAGAGAACTTCGCGCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 788
DB 298 AlaGly-----ProArgPro--GlnValProGlnAspAlaPro 310
OY 789 CAAAGTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 848
DB 311 GlnGlyGly-----AspSerGln 316
OY 849 CGCGCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 908
DB 317 ArgArgGlnThrProProArgProHisSerArg----- 327
OY 909 GCTCAAGCAGGACGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 968
DB 328 -----LysArgArgAspThrGlyAlaHisHisArgHisTrp----- 339
OY 969 CCGCGCATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1028
DB 340 -----ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 348
OY 1029 CGACGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1088
DB 349 ArgGlnGlyAlaLeuProAlaAla-----HisProAspArgArgArgArgArg 365
OY 1089 CGTGCAGCGCATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1148
DB 366 ArgArgAlaHisProAspAla----- 372
OY 1149 CGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1208
DB 373 AlaAlaArgAlaSerValPro--AlaHisAlaProAlaHisArgGlyArgLeuArgAla 391
OY 1209 CGA-----CCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1250
DB 392 ArgGlySerThrAlaAlaValAlaProArgProLeuProArg----- 404

```

```

OY 1231 GACCGCGCTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1310
DB 405 -----GlnHisLeuArgArgArgArgArgGlyArgGlnArgGlyProHis----- 418
OY 1311 CTGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1349
DB 419 --HisArgGlyGlyMetGlnArgGlnProHisAlaGly 430

RESULT 11
VNUA_PRIVKA
ID VNUA_PRIVKA STANDARD; PRT: 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE Probable nuclear antigen
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OX Alphaherpesvirinae; Varicelloviruses.
NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; PubMed=2171211;
RA Vitek C., Kozmik Z., Paces V., Schlim S., Schwyzler M.;
RT "Pseudorabies virus: immediate-early gene overlaps with an oppositely
oriented open reading frame: characterization of their promoter and
enhancer regions.";
RL Virology 179:365-377(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: M3451; AAA7471.1;
DR PIR: B45344; B45344.
FT DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475B95E2 CRC64;

Alignment Scores:
Pred. No.: 3,92e-05
Score: 274.00 Length: 1733
Percent Similarity: 33.43% Matches: 200
Best Local Similarity: 28.94% Mismatches: 229
Query Match: 8.45% Indels: 231
DB: 1 Gaps: 40

US-10-009-782-1 (1-1758) x VNUA_PRIVKA (1-1733)
OY 1735 CTGCGGCTTGGACAGGTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1666
DB 740 ValGlyGlnGlyGlnGlnArgArgGlnAlaAlaAlaAlaAlaAlaAlaAlaAla 754
OY 1665 CATTTTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1639
DB 755 --ValProGlnArgAlaArgGlyAlaLeuGlnGlyLeuGlnGlyAlaGlnLeuValGly 773
OY 1638 CAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1579
DB 774 GlnArgValValGlnHisHisAlaHisValLeuGlnGlyTyrLeuProHisPro 793
OY 1578 GGG-----AGGCGCGCATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1549
DB 794 GlyGlyAlaAlaAlaGlnArgGlyAlaAlaAlaAlaArgGlyAspValArg-----Gln 810

```


RN [2]
 RP SEQUENCE OF 1-181 FROM N.A.
 RX MEDLINE-84270697; PubMed-6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
 RA Myers J., Williams C., Ramirez F.,
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 RT conservation of a pattern of introns and exons.";
 RL Nature 310:337-340(1984).
 RN [3]
 RP SEQUENCE OF 162-301.
 RC TISSUE-SKIN.
 RX MEDLINE-71038625; PubMed-5529814;
 RA Click E.M., Bornstein P.,
 RT "Isolation and characterization of the cyanogen bromide peptides from
 RT the alpha 1 and alpha 2 chains of human skin collagen.";
 RL Biochemistry 9:4699-4706(1970).
 RN [4]
 RP SEQUENCE OF 263-268.
 RC TISSUE-SKIN.
 RX MEDLINE-71001508; PubMed-4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.,
 RT "A comparative study of glycopeptides derived from selected
 RT vertebrate collagens. A possible role of the carbohydrate in fibril
 RT formation.";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [5]
 RP SEQUENCE OF 425-1464 FROM N.A.
 RX MEDLINE-84080385; PubMed-6689127;
 RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.,
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223(1983).
 RN [6]
 RP SEQUENCE OF 1229-1454 FROM N.A.
 RC TISSUE-BONE.
 RX MEDLINE-88124208; PubMed-3340531;
 RA Mackelae J.K., Raasina M., Vaita A., Vuorio E.,
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [7]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE-88097389; PubMed-3480516;
 RA Bonstein P., McKay J., Morishima J.K., Devareyalu S., Gellinas R.E.,
 RT "Regulatory elements in the first intron contribute to
 RT transcriptional control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [8]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE-85130970; PubMed-2857713;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.,
 RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
 RT Promoter structure, Alu repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [9]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE-88033098; PubMed-2822714;
 RA Nosou W.C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.J.,
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-91184577; PubMed-2010058;
 RA Kulvanenti H., Tromp G., Prockop D.J.,
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans.";
 RL FASEB J. 5:2052-2060(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97259599; PubMed-9101290;
 RA Kulvanenti H., Tromp G., Prockop D.J.,
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-91374476; PubMed-1895312;
 RA Byers P.H., Wallis G.A., Walling M.C.,
 RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
 RL J. Med. Genet. 28:433-442(1991).
 RN [13]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97169389; PubMed-9016532;
 RA Dalgleish R.,
 RT "The human type I collagen mutation database.";
 RL Nucleic Acids Res. 25:181-187(1997).
 RN [14]
 RP VARIANT OI-II CYS-1166.
 RX MEDLINE-86287390; PubMed-3016737;
 RA Cohn D.H., Byers P.H., Steinmann B., Gellinas R.E.,
 RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
 RT change in one human pro alpha 1(I) collagen allele.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
 RN [15]
 RP VARIANT OI-II ARG-569.
 RX MEDLINE-87222295; PubMed-3108247;
 RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.,
 RT "A point mutation in a type I procollagen gene converts glycine 748
 RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
 RT a lethal variant of osteogenesis imperfecta.";
 RL J. Biol. Chem. 262:14737-14744(1987).
 RN [16]
 RP VARIANT OI-II CYS-926.
 RX MEDLINE-88033031; PubMed-3667599;
 RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.,
 RT "A point mutation in a type I procollagen gene converts glycine 748
 RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
 RT a lethal variant of osteogenesis imperfecta.";
 RL J. Biol. Chem. 262:14737-14744(1987).
 RN [17]
 RP VARIANT OI-II ARG-842.
 RX MEDLINE-88298828; PubMed-3403550;
 RA Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.,
 RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
 RT chain in lethal perinatal osteogenesis imperfecta. Demonstration of
 RT the peptide defect by in vitro expression of the mutant cDNA.";
 RL J. Biol. Chem. 263:11627-11630(1988).
 RN [18]
 RP VARIANT OI CYS-1195.
 RX MEDLINE-89218628; PubMed-3244312;
 RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.,
 RT "A cysteine for glycine substitution at position 1017 in an alpha
 RT 1(I) chain of type I collagen in a patient with mild dominantly
 RT inherited osteogenesis imperfecta.";
 RL Mol. Biol. Med. 5:197-207(1988).
 RN [19]
 RP VARIANT OI-II VAL-434.
 RX MEDLINE-89255493; PubMed-2470760;
 RA Patterson E., Smiley E., Bonadio J.,
 RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
 RT mutation.";
 RL J. Biol. Chem. 264:10083-10087(1989).
 RN [20]
 RP VARIANT OI-IV SER-1010.
 RX MEDLINE-89308591; PubMed-2745420;
 RA Matini J.C., Grange D.R., Gottesman G.S., Lewis M.B., Koepflin D.A.,
 RT "Osteogenesis imperfecta type IV. Detection of a point mutation in
 RT the alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
 RL J. Biol. Chem. 264:11893-11900(1989).
 RN [21]
 RP VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.

Alignment Scores:	
Pred. No.:	5.85e-05
Score:	270.00
Percent Similarity:	31.958
Best Local Similarity:	26.788
Query Match:	8.108
DB:	1
Length:	1664
Matches:	181
Conservative:	35
Mismatches:	269
Indels:	152
Gaps:	33

OY	456	CAACGGCGCTGTAT-----		47
Db	491	GIYsetArgIyPheProGlyAlaAspIyValAlaGlyProLysGlyProAlaGlyIu		51
OY	471	-----	GGGGCGCATTCACCGCTGGCGCGCGGT	50
Db	511	ArgGlysetProGlyProAlaGlyProLysGlySerProGlyGluAlaIyArProGly		53
OY	501	CATGGCGGACTTCAGCGCGCGCGCCACCGAGAGAAATCGCGCATCGGAGACTGC		56
Db	531	GluAlaAGlyLeuProGlyAlaIyAsGlyLeuThrGlySerProGlySerProIyPro		55
OY	561	CGAGGAAGCATGCGCGAGCGCGCGCATTCGGACCGCGCGCTCTACCCCGCGC		62
Db	550	spGlyIyThrGlyProProGlyProProGlyIyGluAspIyArProGly-----		56
OY	621	CGCCCGCGCCACCA-----	CCGAAGA	64
Db	567	roProGlyIyProProGlyIyAlaArGlyIyGluAlaGlyIyAlaMetIyPheProGlyProLysG		58
OY	642	GATCATGAGAGTGTGCCCGCGCGCTGAGCGCGCATGGCGCGCATCTACGCCACCATCG		70
Db	587	IyAlaAlaGlyIyIuProGlyIyIyAlaGlyIyIuArGlyIyAlaProGlyProProGlyIyAlaV		60
OY	702	CGAGGAAGCGAGACACATCGTGTGGCGCGCGTGAGGAACCTTCGCCATCGCGCGGACT		76
Db	607	AlGlyProAlaGlyIyIyAsp-GlyIyAlaAlaGlyIyAlaGlyIyProProGlyPro--		Ala 625
OY	762	GGAGCGCGCGGT-----	GCTGATCTCGCACCAAGGTCAI-----	797
Db	626	GlyProAlaGlyIyGluArGlyIyGluGlyIyProAlaGlySerProGlyIyPheGlyIyLeu		645
OY	798	---GGGCGAGCCCAATTTGGCGCGCTCGCGCGAGACGCTGGCGCGTATCAGAGCGCCCAT		854
Db	646	ProGlyIyProAlaGly-----	ProProGlyIyGluAlaGlyIyIyProGlyIyGluIu	661
OY	855	GGGCGCGCGAGACGTCTGCT-----	GGACGCGTATCCCTACGCGCGCGGTCCACCAT	908
Db	662	GlyValProGlyIyAspLeuGlyIyAlaProGlyProSerGlyIyAlaArGlyIyIuArGlyIyPhe		661
OY	909	-----	GCTACAGCAGACCGCGCTGCTGGCGCG	938
Db	682	ProGlyIyIuArGlyIyValGlyIyIuGlyIyProProGlyProAlaGlyProAlaGlyIyAlaAsn		GI 701
OY	939	AGGACACATCATCAGCTGTCGCAAGCCCTTCCGAACTGAGCGCGCGCATCGTATGA		998
Db	701	IyAlaProGlyIyAsnAspIyIyAlaIyGlyIyAspAlaGlyIyAlaProGlyIyAla-ProGly---		719
OY	999	AGTCGCGCGCGAGCGCGCAAAATCAGACTACAGCTGTGCTCCCGACCTCGACCGCGCGG		1056
Db	720	SerGlyIyIyAlaProGlyIyLeuGlyIyMetProGlyIyIuArGlyIyIyAlaIyGlyIyLeu		738
OY	1059	CGCCATCTACTTCATGATGACGAAACCGAGCTGACGCGCATTCCTGCTGGCGCGAC		1118
Db	739	-----	ProGlyIyProLysIyAsp	744
OY	1119	CATGATCGGCTCGAGGCGCTGCCCGCAGCAGCGCCCGCA-----		1160
Db	745	ArgGlyIyAspAlaGlyProLysGlyIyAlaAspIySerProGlyIyIyAspIyIyAlaIyArGlyIy		764
OY	1161	---TCGCGCGCTGTGGGCGACCTT-----	CCGCGCGGCGTGGGGCACTATGCCGCGGACCT	1214
Db	765	LeuThrGlyIyProIyIyGlyProProGlyProAlaGlyIyAlaProGlyIyAspIyIyGlyIySer		784
OY	1215	GGGCGCTT-----	CCGCTGGAAGACGGCGGTANGGAAGATGACCGCGCTGA-	1261
Db	785	GlyProSerGlyProAlaGlyProThrGlyIyAlaArGlyIyAlaProGlyIyAspArGlyIyIu		804
OY	1262	-----	CCGCGCGCGCGCTTGGCGCTGGCGCGGCG	1288
Db	805	ProGlyIyProProGlyProAlaGlyIyPheAlaGlyIyProProGlyIyAlaAspIyGlyIuProGlyIy		824


```

CC CC      RESULT 14
CC CC      CALL_CANPA
CC CC      ID          CALL_CANPA
CC CC      AC          O9XSUT;
CC CC      DT          30-MAY-2000 (Rel. 39, Created)
CC CC      DT          30-MAY-2000 (Rel. 39, Last sequence update)
CC CC      DT          30-MAY-2000 (Rel. 39, Last annotation update)
CC CC      DE          Collagen alpha 1(I) chain precursor.
CC CC      GN          COL1A1.
CC OS       Canis familiaris (Dog).
CC OC       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC       Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
CC OC       NCBI_TaxID=9615;
CC RN       [1]
CC RP       SEQUENCE FROM N.A.
CC RC       TISSUE-SKIN;
CC RA       Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
CC RT       "Sequence of normal canine COL1A1 cDNA.";
CC RL       Submitted (MAY-1999) TO THE EMBL/GENBANK/DDJJ databases.
CC CC      - FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC CC      (FIBRILLAR FORMING COLLAGEN).
CC CC      - SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC CC      - UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC CC      - SIMILARITY: CONTAINS 1 WFCC DOMAIN.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----

```

```
DR EMBL; AF153062; AAD34619.1;
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
```

Alignment Scores:	
Pred. No.:	6 76e-05
Score:	266.50
Percent Similarity:	31.92%
Best Local Similarity:	27.70%
Query Match:	8.14%
DB:	1
	C
Length:	1460
Matches:	177
Conservative:	27
Mismatches:	27
Indels:	188

Gap: 29

QY	106	CCGGGGGGGGCCCGGACCTGGGCGCTGGCGGGCGGACCGGATCGCGC	15
Db	435	ProGlyAsnLysPheGlyAspThrGlyAlaLysGlu- ProGlyProThrGlyIleGlnI	45
QY	152	-----CCATCGCGCATCTGTGTGGAGCGCGCGCCGACAC-----	18
Db	454	yProProGlyProAlaGly- GluGlnGlyLysAspGlyAlaAspGlyLysIubProGlyProT	474
QY	186	-----CGGGGTGCAGCTGTGTGGCTCGGTGGTGTGGCGCGCGCTTCATC	228
Db	474	hGlyLysuProGlyProProGlyGlyLysGlyGlyProGlySerThrGly-	490
QY	229	GACTCGCACACCCAGCAGCAGCAACATCTGCTCAGCGCTCGCATATGACGCCCAAGATC	288
Db	490	-----	490
QY	289	TCCGAGGGCGGTACACACAGGTGTACAGGGCAATTGGCGCATACGCTTGGCGCGCTGGCG	348
Db	491	-----PheProGlyAlaAspGlyY	497
QY	349	CACGCCAACCGCGCGCGCGCGCGCTGTGACATGCTGTGGACAGAGCGGCTTACGGTTTCGAG	408
Db	497	al-----AlaIlyProLysGlyProAlaGlyLysY	507
QY	409	CGCTTCGCCGACTACCTGAGCGCTTGGGAGCCAGCGCGCGCGCTCAACCGCGCTGT	468
Db	508	-----GlySerProGlyProAlaGlyProLys	516
QY	469	ATGTGGGGCCATTCAAGCGTGTGGCGCGCGCGGTATCCGAGATTTCAGAGCGCGCCACC	528
Db	517	--GlySerProGlyAlaAlaGlyYAspProGlyLysIubAlaLysIleuProGlyAlaLysGlyL	536
QY	539	GACGAGGAATTCGGCGCATGTGGGAGACCTGGCGCGAGAAACCATGGCCAGCGCGCCATC	588
Db	536	euthrGlySerProGlySerProGlyPro-AspLysYThrGlyProProGlyProAl	555
QY	589	GGCATTTTCGACCGCGCGCTTTCATCCGCGCGCGCGCGCGCGCACCA	634
Db	555	agLysLAspGlyYAspProGlyY-----ProProGlyProProGlyAlaAspGlyGly	572

```

Db      903  ysglyalaarglygluthtgilyproglinaaglyargproglinaaglyvalaglyleuproproglyp 923
Oy      1617  -----GGAAGTAGTAGGGGGCGCGCGCTCGCGCCCGCAGCCGCAAGAGCCCGGAAATGACCTG 1671
Db      923  roproglyproalaaglyalutysglsyserproglyalaaspglyproalaaglyalaarg 943
Oy      1672  CAAGACCT-----GTGCGAGCCAGCGG-----CATC 1658
Db      943  lythproglyproglinaaglyalaaglyalaaglyalaaglyalaaglyalaaglyleuproglylna 963
Oy      1699  GCGGCTCGACCCCTCTCCAGGCCGAGCTGGGCCA 1733
Db      963  rgglygluarglypneuproglyleuproglypro 974

RESULT 15
SE5_RAT
AC      5E5_RAT      STANDARD;      PRT;      825 AA.
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      5E5 antigen.
OS      Rattus norvegicus (Rat.).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Mistar; TISSUE=Brain;
RC      MEDLINE=96015159; PubMed=8537300;
RA      Suzuki E., Kojima N., Yoshimura K., Dyemura K., Obata K., Akagawa K.;
RT      "Cloning and sequence analysis of cDNA for a possible DNA-binding
RL      protein 5E5 in the nervous system.";
J. Biochem. 118:122-128(1995).
CC      -1- FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -1- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; D37934; BAA07153.1;
KM      DNA-binding; Nuclear protein; Antigen.
SQ      SEQUENCE 825 AA; 86831 MW; AFE67FE2FD555BDF CRC64;

Alignment Scores:
Pred. No.:      8.22e-05      Length:      825
Score:      267.00      Matches:      158
Percent Similarity:      32.33%      Conservative:      36
Best Local Similarity:      26.33%      Mismatches:      221
Query Match:      8.09%      Indels:      185
Db:      1      Gaps:      28

US-10-009-782-1 (1-1758) x SE5_RAT (1-825)

Oy      3  ATTCCACTGATTCGCGGAGAGAGATTTCATGTCCTCCAAATCCGATTCAGACCCCTTGA 62
Db      315  ValProlysaPaPaGlygluL----- 321
Oy      63  CCTCTGCTGCTCGCGGCGGACCCATCATGACGAGACAAACCCCGGCGGCGCGCGGA 122
Db      322  ---GlyGlyArgGlyLTPpGlyProGlyAlaAlaGlnGlnGlyGlyAlaAlaArg-As 340
Oy      123  CTGAGCGGTGCGCGGCGGACGACGATCGCGGCGGATCTGTCGAGACGCGCGCGGA 182
Db      340  p-TriphtSerGluSer-----ProArgn 348

```


GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 01:04:21 ; Search time 4570 Seconds

(without alignments)
1195.346 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 1758

Sequence: 1 gaattccactgacgagga.....ccctgagctacgagaagctt 1758

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database: Listing first 45 summaries

GenBank:*

1: gb_ba:*

2: gb_hg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_ov:*

7: gb_ov:*

8: gb_ov:*

9: gb_ov:*

10: gb_ov:*

11: gb_ov:*

12: gb_ov:*

13: gb_ov:*

14: gb_ov:*

15: gb_ov:*

16: gb_ov:*

17: gb_ov:*

18: gb_ov:*

19: gb_ov:*

20: gb_ov:*

21: gb_ov:*

22: gb_ov:*

23: gb_ov:*

24: gb_ov:*

25: gb_ov:*

26: gb_ov:*

27: gb_ov:*

28: gb_ov:*

29: gb_ov:*

30: gb_ov:*

31: gb_ov:*

32: gb_ov:*

33: gb_ov:*

34: gb_ov:*

35: gb_ov:*

36: gb_ov:*

37: gb_ov:*

38: gb_ov:*

39: gb_ov:*

40: gb_ov:*

41: gb_ov:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1758	100.0	1758	BD011857	BD011857 Transform
2	1758	100.0	1758	E59056	E59056 Transformed
3	1758	100.0	1758	BD0004254	BD0004254 Transformed
4	1739.4	98.9	1980	S80683	S80683 dan-D-amin
5	1433.4	82.7	1455	AFANADA	AF333548 Alcaligen
6	1088.6	61.9	1455	AF332548	AF333548 Alcaligen
7	693.6	39.5	1497	AFPA61	D45919 Alcaligenes
8	533.6	30.4	208050	AL646083	AL646083 Ralstonia
9	455.8	25.9	1467	AFPA6	D50061 Alcaligenes
10	306	17.4	110000	LMFLCHR34_10	Continuation (11 o
11	171.8	9.8	22251	SC2K36	Continuation (11 o
12	116	6.6	82746	AF453501	AF453501 Actinomy
13	107.4	6.1	28751	SC6A11	AF453501 Actinomy
14	99.2	5.6	14710	AE006004	AE006004 Caulobact
15	99.2	5.6	123580	AX211705	AX211705 Sequence
16	99.2	5.6	123580	AF263912	AF263912 Streptomy
17	99.2	5.6	123580	AX211739	AX211739 Sequence
18	97.2	5.5	113193	AF357202	AF357202 Streptomy
19	94.2	5.4	5123	HEPVE	AF357202 Streptomy
20	94.2	5.4	8438	SHILT	X15120 Pseudorab
21	94	5.3	4039	AF425991	M57505 Pseudorab
22	94	5.3	34766	SCC121	AF425991 Streptomy
23	93.6	5.3	110000	LMFLCHR32_06	Continuation (7 of
24	93.4	5.3	27541	AX211706	AX211706 Sequence
25	91.4	5.2	37501	SC6D7	AL133213 Streptomy
26	91	5.2	35033	SC7H1	AL096811 Streptomy
27	91	5.2	35654	SC7H1	AL021411 Streptomy
28	89.6	5.1	15141	SHIPROIE	M34651 Pseudorab
29	89.6	5.1	33779	SCGD3	AL096822 Streptomy
30	88.8	5.1	110000	LMFLCHR36_31	Continuation (32 o
31	88.2	5.0	38404	SC2G5	AL035478 Streptomy
32	88	5.0	13848	AE005104	AE005104 Halobacte
33	87.8	5.0	1746	AF425992	AF425992 Streptomy
34	87.8	5.0	30657	SCD12A	AL357524 Streptomy
35	86.8	4.9	41906	SC5C7	AL031515 Streptomy
36	86	4.9	88421	AX417445	AX417445 Sequence
37	85.6	4.9	5520	AC084329	AC084329 Leishman
38	85.6	4.9	49736	AF319998	AF319998 Streptomy
39	85.2	4.8	25362	SCF56	AF319998 Streptomy
40	84.6	4.8	109519	AX195929	AX195929 Sequence
41	84	4.8	1839	AX196078	AX196078 Sequence
42	83.8	4.8	3578	AF005277	AF005277 Cellulono
43	83.8	4.8	3619	SCVALSFP	Y13070 S.coelicolo
44	83.8	4.8	24800	SC9B1	AL049727 Streptomy
45	83.8	4.8	110000	LMFLCHR36_03	Continuation (4 of

ALIGNMENTS

RESULT 1

BD011857

LOCUS BD011857 1758 bp DNA linear PAT 02-AUG-2002

DEFINITION Transformant microorganism and process for producing

ACCESION BD011857

VERSION BD011857

KEYWORDS WO 0078926-A/1.

SOURCE UNIDENTIFIED

ORGANISM UNIDENTIFIED

REFERENCE 1 (bases 1 to 1758)

AUTHORS Takeuchi, K., Koida, Y., Hirose, Y., Moriwuchi, M. and Isobe, K.

TITLE Transformant microorganism and process for producing D-aminocyclase

JOURNAL Patent: WO 0078926-A 1 28-DEC-2000,

Pred. No. is the number of results predicted by chance to have a

COMMENT

AMANO PHARMACEUTICAL CO LTD, KENICHI TAKEUCHI, YOSHINAO KOIDE,
YOSHIHIKO HIROSE, MITSUKI MORIGUCHI, KIMIYASU ISOBE
OS ALCALIGENES xylosoxydans subsp. xylosoxydans
PN MO 0078926-A/1
PD 28-DEC-2000
PF 15-JUN-2000 MO 2000JP003932
PR 17-JUN-1999 JP 99P 170555
PI KENICHI TAKEUCHI, YOSHINAO KOIDE, YOSHIHIKO HIROSE, PI
MITSUKI MORIGUCHI,
KIMIYASU ISOBE
PC C12N1/21, C12N15/52, C12N9/80
CC
FH

FEATURES

Source Location/Qualifiers
1. 1758
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 281 a 662 c 568 g 247 t
ORIGIN

Query Match 100.0%; Score 1758; DB 6; Length 1758;
Best Local Similarity 100.0%; Pred. No. 9e-183;
Matches 1758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GAATTCACCTTATCGCGGAGAGATTTCCATGTCCCAATCCGATCCAGGCTTC 60
DB 1 GAATTCACCTTATCGCGGAGAGATTTCCATGTCCCAATCCGATCCAGGCTTC 60
QY 61 GACCTGCTCTCGCGGGGCGACCTTCATCGAGGAGACAAACCCCGGGGCGCGC 120
DB 61 GACCTGCTCTCGCGGGGCGACCTTCATCGAGGAGACAAACCCCGGGGCGCGC 120
QY 121 GACCTGGGCGTCCGGGGGAGCCGATCGCCGCGATCGGCGATGTGTGGAGCGCGCG 180
DB 121 GACCTGGGCGTCCGGGGGAGCCGATCGCCGCGATCGGCGATGTGTGGAGCGCGCG 180
QY 181 CACACCCGGGTGCACTGTGTGGGCTGTGTGTGGGCGGCTTCATCTGACTGCGACAC 240
DB 181 CACACCCGGGTGCACTGTGTGGGCTGTGTGTGGGCGGCTTCATCTGACTGCGACAC 240
QY 241 CACGACACAACCTACCTGCTCAGGCGTGGCGATGACGCGCAAGATCTGCGAGGCGTC 300
DB 241 CACGACACAACCTACCTGCTCAGGCGTGGCGATGACGCGCAAGATCTGCGAGGCGTC 300
QY 301 ACCACGGGTGTACAGGGCAATTCGGGATCAGGCTGGCGCGTGGCGCAACCGG 360
DB 301 ACCACGGGTGTACAGGGCAATTCGGGATCAGGCTGGCGCGTGGCGCAACCGG 360
QY 361 CCCGCCCTTGACCTGCTGAGCAAGCGGCTTTACGTTTGAAGGCTTCGCGAC 420
DB 361 CCCGCCCTTGACCTGCTGAGCAAGCGGCTTTACGTTTGAAGGCTTCGCGAC 420
QY 421 TACCTGGACGCTTGGGGGCGACGCGGCGGCGTCAACGCGCGCTGTATGTGGCCAT 480
DB 421 TACCTGGACGCTTGGGGGCGACGCGGCGGCGTCAACGCGCGCTGTATGTGGCCAT 480
QY 481 TCAAGCTGCGCGCGCGGTATGCGGCACTTGCAGGGCGCGCGACCGAGAGAAATC 540
DB 481 TCAAGCTGCGCGCGCGGTATGCGGCACTTGCAGGGCGCGCGACCGAGAGAAATC 540
QY 541 GCGGCGATGCGGAGCTTGGCGGAGAGACATGCGCAGCGCGCGCATCGGCAATTCGAC 600
DB 541 GCGGCGATGCGGAGCTTGGCGGAGAGACATGCGCAGCGCGCGCATCGGCAATTCGAC 600
QY 601 GCGGCGCTTACCGCGCGCGCGCGCGCAACCGMAAGATCATGAGGTGTGCGG 660
DB 601 GCGGCGCTTACCGCGCGCGCGCGCGCAACCGMAAGATCATGAGGTGTGCGG 660
QY 661 CCGGTGAGCGGATGCGGCGCATCTAGCCACCATGCGGAGAGAGCGGCGACATC 720
DB 661 CCGGTGAGCGGATGCGGCGCATCTAGCCACCATGCGGAGAGAGCGGCGACATC 720
QY 721 GTGGCGCGCTGAGGAAACCTTCGCGATCGCGCGACGTGACGTGGGTGATC 780

```

```

DB 721 GTGGCGCGCTGAGGAAACCTTCGCGATCGCGCGACGTGAGCGGTGTATC 780
QY 781 TCGACACCAAGATCATGCGGCGACCCCAATTCGCGCGCTGCGGCGAGAGCGTCCGCTG 840
DB 781 TCGACACCAAGATCATGCGGCGACCCCAATTCGCGCGCTGCGGCGAGAGCGTCCGCTG 840
QY 841 ATGAGGCGCGCATGCGGCGCGAGAGCTGCTGCGTGAAGCGGTATCCCTAGTGGCGGCG 900
DB 841 ATGAGGCGCGCATGCGGCGCGAGAGCTGCTGCGTGAAGCGGTATCCCTAGTGGCGGCG 900
QY 901 TCCACCATGCTCAAGCAGACCGCGTGTGCTGCGGCGAGCGACCATCATACCTGTGTC 960
DB 901 TCCACCATGCTCAAGCAGACCGCGTGTGCTGCGGCGAGCGACCATCATACCTGTGTC 960
QY 961 AAGCCCTTCCCGAATGAGCGGCGCGCACTGATGAGTGCAGCGCGCGCGGCGGCAAA 1020
DB 961 AAGCCCTTCCCGAATGAGCGGCGCGCACTGATGAGTGCAGCGCGCGCGGCGGCGGCAAA 1020
QY 1021 TCCAAATGACAGTGTGCGCGAGCTGCGAGCGCGCGCGCGCGCATCTTATGATGAGAC 1080
DB 1021 TCCAAATGACAGTGTGCGCGAGCTGCGAGCGCGCGCGCGCGCATCTTATGATGAGAC 1080
QY 1081 GAACCCGACGTGCAAGCGCATCTGCGGCTTGGCGCGACCATGATCGGCTCGAGCGCTG 1140
DB 1081 GAACCCGACGTGCAAGCGCATCTGCGGCTTGGCGCGACCATGATCGGCTCGAGCGCTG 1140
QY 1141 CCGACAGCAGAGCGCGCGCATCGCGGCTGTGTGGGCACTTCCCGCGGCTGTGGGCGAC 1200
DB 1141 CCGACAGCAGAGCGCGCGCATCGCGGCTGTGTGGGCACTTCCCGCGGCTGTGGGCGAC 1200
QY 1201 TATGCGGCGGACCTGTGGCTTCCGCTGAGAGCGCGGTATGAGAAATGACCGGCTG 1260
DB 1201 TATGCGGCGGACCTGTGGCTTCCGCTGAGAGCGCGGTATGAGAAATGACCGGCTG 1260
QY 1261 ACCGCGCGGCTTGGCGGCGGCGGCGGCGGCGAGCTGCAAGCGCGGTACTTCCGCGAC 1320
DB 1261 ACCGCGCGGCTTGGCGGCGGCGGCGGCGGCGAGCTGCAAGCGCGGTACTTCCGCGAC 1320
QY 1321 CTGGGTGTGTGACCGCGGCGACGCGGTGCGATACCGGCACTTGCATAACCTTACCGAG 1380
DB 1321 CTGGGTGTGTGACCGCGGCGACGCGGTGCGATACCGGCACTTGCATAACCTTACCGAG 1380
QY 1381 CCGCGCGCGGCGATCATCTCGGTACGTAACGCGCGCGCGGTGTGCAAGAGAGGCG 1440
DB 1381 CCGCGCGCGGCGATCATCTCGGTACGTAACGCGCGCGGTGTGCAAGAGAGGCG 1440
QY 1441 TTACCGGCGACAGATCGCGCGCGGTGTGCGAGCGAGCGCGCTGAGCGCGCGCAG 1500
DB 1441 TTACCGGCGACAGATCGCGCGCGGTGTGCGAGCGAGCGCGCTGAGCGCGCGCAG 1500
QY 1501 CCTTACAAATCGGCGGTGACGCGGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGAG 1560
DB 1501 CCTTACAAATCGGCGGTGACGCGGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGAG 1560
QY 1561 CGCTACATGCGCGCTTCCCTCGGCTCGCATACGCGCGCGCGCGCGCGCGCGCGCGAG 1620
DB 1561 CGCTACATGCGCGCTTCCCTCGGCTCGCATACGCGCGCGCGCGCGCGCGCGCGCGAG 1620
QY 1621 GTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1680
DB 1621 GTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1680
QY 1681 TCGCAGCGCGAGCGCATTCGCGGTGTGCAACCGCGCGCGCGCGCGCGCGCGCGAG 1740
DB 1681 TCGCAGCGCGAGCGCATTCGCGGTGTGCAACCGCGCGCGCGCGCGCGCGCGCGAG 1740
QY 1741 CTGAGCTACGAAAGCTT 1758
DB 1741 CTGAGCTACGAAAGCTT 1758

```

RESULT 2

[illegible]

Db	61	GACCTGTGTCGGCGGGGCGACACCTTCATCGACGGCAGCACACCCGGGGGCGCGCC	120
Qy	121	GACCTGGGCGTGGCGGGCGACCGCATCGCGGCATGGCGATCTGTGGAGCGCGCGCG	180
Db	121	GACCTGGGCGTGGCGGGCGACCGCATCGCGGCATCGCGCATCTGTGGAGCGCGCGCG	180
Qy	181	CACACCGGGGTGCAGCTGTGCGGGCGGTGGGTGGCGCGCGCGCTTCATCGACTGCGACACC	240
Db	181	CACACCGGGGTGCAGCTGTGCGGGCGGTGGGTGGCGCGCGCGCTTCATCGACTGCGACACC	240
Qy	241	CACAGACGACAATCTACTGTCTAGGGGTGCGCGACATGAGCGCCCAAGATCTGCGAGGGCGTC	300
Db	241	CACAGAGACAATCTACTGTCTAGGGGTGCGCGACATGAGCGCCCAAGATCTGCGAGGGCGTC	300
Qy	301	ACCAGGCTGTACCGGGGCAATTGGCGGCATCAGCCTGGCGCGCGCTGGCGCACGCCAACCG	360
Db	301	ACCAGGCTGTACCGGGGCAATTGGCGGCATCAGCCTGGCGCGCGCTGGCGCACGCCAACCG	360
Qy	361	CCCGCCCCCTGGAGCTGTGTGACACAAAGCGGCTTACGGTTTCAGAGGCTTGGCGAC	420
Db	361	CCCGCCCCCTGGAGCTGTGTGACACAAAGCGGCTTACGGTTTCAGAGGCTTGGCGAC	420
Qy	421	TACTTGGAGGCGCTTGCGGGCCACGCGCGCGCGCTCAAGCGCGCTGTATGATGGGCAT	480
Db	421	TACTTGGAGGCGCTTGCGGGCCACGCGCGCGCGCTCAAGCGCGCTGTATGATGGGCAT	480
Qy	481	TCAAAGCTTGGCGCGCGGTATGCGCGATCTTGAAGGCGCGCGCACCGGACGAGGAAATC	540
Db	481	TCAAAGCTTGGCGCGCGGTATGCGCGATCTTGAAGGCGCGCGCACCGGACGAGGAAATC	540
Qy	541	GGGGGCATGGCGGACCTGGCGGAGAAACCATGGCCAGCGCGGCATCGGCATTTTCAGCC	600
Db	541	GGGGGCATGGCGGACCTGGCGGAGAAACCATGGCCAGCGCGGCATCGGCATTTTCAGCC	600
Qy	601	GGCGCTTCTACCGCGCGCGCGCGCGCGGACACACGAAAGATCATCGAGGTGGCGG	660
Db	601	GGCGCTTCTACCGCGCGCGCGCGCGCGGACACACGAAAGATCATCGAGGTGGCGG	660
Qy	661	CCGCGAGCGGGGCATGGGGGATCTACGCCACCCACATGGCGGACGAGCGAGCAGCATC	720
Db	661	CCGCGAGCGGGGCATGGGGGATCTACGCCACCCACATGGCGGACGAGCGAGCAGCATC	720
Qy	721	GTGGCGCGGTGGAGGAAACCTTCGCGCATGGCGGGAGCGTGGAGTGGCGGTGGTATC	780
Db	721	GTGGCGCGGTGGAGGAAACCTTCGCGCATGGCGGGAGCGTGGAGTGGCGGTGGTATC	780
Qy	781	TCGCAACCAAGGTCTATGGGCGACGCCAATTTTCGCGCGCTCCGCGAGACGCTGGCGGTG	840
Db	781	TCGCAACCAAGGTCTATGGGCGACGCCAATTTTCGCGCGCTCCGCGAGACGCTGGCGGTG	840
Qy	841	ATCGAGGCGCGCATGGGCGCGCGCGACGATCTGCTGTGAGCGGTATCCTTACGTGGCGGC	900
Db	841	ATCGAGGCGCGCATGGGCGCGCGCGACGATCTGCTGTGAGCGGTATCCTTACGTGGCGGC	900
Qy	901	TCGACATGCTCAAGCAGGACCGCGTGTGCTGGCGCGGACGACCATATCACTGGGTGC	960
Db	901	TCGACATGCTCAAGCAGGACCGCGTGTGCTGGCGCGGACGACCATATCACTGGGTGC	960
Qy	961	AAGCGCTTCCCGCAACTGAGCGGGCGCGACCTGGATGGAATGCGCGCGCGAGCGCGGCAAA	1020
Db	961	AAGCGCTTCCCGCAACTGAGCGGGCGCGACCTGGATGGAATGCGCGCGCGAGCGCGGCAAA	1020
Qy	1021	TCGCAAGTACAGCTGTGGCGCGGAGGTGAGCGCGCGCGCGCGCATCTACTTCATGATGAC	1080
Db	1021	TCGCAAGTACAGCTGTGGCGCGGAGGTGAGCGCGCGCGCGCGCATCTACTTCATGATGAC	1080
Qy	1081	GAAACCGAGCTGACGCGCATCTCTGGCGTTTCGGCGCGGACATGATCGGCTCCGAGCGGCTG	1140
Db	1081	GAAACCGAGCTGACGCGCATCTCTGGCGTTTCGGCGCGGACATGATCGGCTCCGAGCGGCTG	1140
Qy	1141	CGCGAGAGAGCGCCCGCATCGCGCGCTGTGGGGCACCTTCCGCGGGGTGCGGGGAC	1200
Db	1141	CGCGAGAGAGCGCCCGCATCGCGCGCTGTGGGGCACCTTCCGCGGGGTGCGGGGAC	1200

Db 1141 CCGCAGACGAGAGCCCGCATCCGCGCTGTGGGCACTTCCCGGGGTGCTGGGCGAC 1200

Qy 1201 TATGGCGCGGACCTGGGCGCTTTCCCGCTGGAGAGCGCGGTATGAAATGACCGGCTG 1260

Db 1201 TATGGCGCGGACCTGGGCGCTTTCCCGCTGGAGAGCGCGGTATGAAATGACCGGCTG 1260

Qy 1261 ACCGCGCGCGCTTGGGCGCTGGCGCGCGCGAGCTCAAGCGCGGTACTTCCCGGAC 1320

Db 1261 ACCGCGCGCGCTTGGGCGCTGGCGCGCGAGCTCAAGCGCGGTACTTCCCGGAC 1320

Qy 1321 CTGGTGGGTTCGACCGCGCGCGAGTGGCGCATACCGCGCATCTTGAACACCTTACCGAG 1380

Db 1321 CTGGTGGGTTCGACCGCGCGCGAGTGGCGCATACCGCGCATCTTGAACACCTTACCGAG 1380

Qy 1381 CGCGCGCGCGCATCTTCCGCTGATGCAACGCGCGCGCGGTTCGCAAGAGAGCGCG 1440

Db 1381 CGCGCGCGCGCATCTTCCGCTGATGCAACGCGCGCGCGGTTCGCAAGAGAGCGCG 1440

Qy 1441 TTCACCGCGCGCATGCGCGCGCGCTGCTGCAAGCGCGCGCGCTGAGCGCGCGCGAG 1500

Db 1441 TTCACCGCGCGCATGCGCGCGCGCTGCTGCAAGCGCGCGCGCTGAGCGCGCGCGAG 1500

Qy 1501 CCCTTACATCCGCGCGGTAAAGCGCGCGCGCTGCGCCCTCCCAACCTTGAAGCAAAAC 1560

Db 1501 CCCTTACATCCGCGCGGTAAAGCGCGCGCGCTGCGCCCTCCCAACCTTGAAGCAAAAC 1560

Qy 1561 CGCTACATGCGCGCGCTGCTGCAATGAGCGCGCGCGCGCTGATGCTGGCAAGAA 1620

Db 1561 CGCTACATGCGCGCGCTGCTGCAATGAGCGCGCGCGCGCTGATGCTGGCAAGAA 1620

Qy 1621 GTGATGGCGCGCGCGCTGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680

Db 1621 GTGATGGCGCGCGCGCGCTGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680

Qy 1681 TCGCAGCGCGCGCGCGCGCGGTCTGCAACCTGTCAGAGCGCGCGCGCGCGCGCGCG 1740

Db 1681 TCGCAGCGCGCGCGCGCGCGGTCTGCAACCTGTCAGAGCGCGCGCGCGCGCGCGCG 1740

Qy 1741 CTGAGCTACGAGAGCTT 1758

Db 1741 CTGAGCTACGAGAGCTT 1758

RESULT 4
S80683 1980 bp DNA linear BCT 13-AUG-1996
LOCUS dan-D-aminocyclase [Alcaligenes xylosoxydans, ssp. xylosoxydans,
A-6, Genomic, 1980 nt].
ACCESSION S80683
VERSION S80683.1 GI:1488356
KEYWORDS
SOURCE Achromobacter xylosoxydans A-6 ssp. xylosoxydans.
ORGANISM Achromobacter xylosoxydans
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
Achromobacter.
REFERENCE 1 (bases 1 to 1980)
AUTHORS Wakayama, M., Katsumo, Y., Hayashi, S., Miyamoto, Y., Sakai, K. and
Moriguchi, M.
TITLE Cloning and sequencing of a gene encoding D-aminocyclase from
Alcaligenes xylosoxydans subsp. xylosoxydans A-6 and expression of
the gene in Escherichia coli
JOURNAL Biosci. Biotechnol. Biochem. 59 (11), 2115-2119 (1995)
MEDLINE 96100942
PubMed 8541651
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gi259174864] from the original journal article.
This sequence comes from Fig. 3.
FEATURES
source 1. 1980
Location/Qualifiers
/organism="Achromobacter xylosoxydans"
/db_xref="taxon:85698"
201..1655
gene /gene="dan"

CDS

/note="D-aminocyclase, N-acyl-D-amino acid amidohydrolase"
201..1655
/gene="dan"
/note="3.5.1.14; Method: conceptual translation with
partial peptide sequencing; This sequence comes from Fig.
3; N-acyl-D-amino acid amidohydrolase"
/codon_start=1
/transl_table=11
/product="D-aminocyclase"
/protein_id="AAB3581.1"
/db_xref="GI:1488357"
/translation="MSQSDSPEDLLAGLLIDSSNTGRRADLVRDRRAATGDL
SDAAATRVDSGLVYAPGFDISHDDNYLLRRMTKRIQGTVTGKGISLA
PLAHPAPADLLDGGSTYRPERFADYDALRIIPAAVNAACVGHSTLRANPDL
ORATDEPILAMDLAEBAASGAIISITGATPPAPAAATPEEIIIEVCRPLSAGIY
ATMRBGEITVALETRIGELDLPVYIHHKVMGPNGRRETLPLEAMAR
OVDSIDAPVPAASITMLKODRYLAFRTIITKCPPELSGDLDEVAERKSTYD
VPELOPAGAVIEMDEPDVQRIALFPTVIGSDGLPHDERPRLMGTPPRLTGYAR
DLGLPFLTAIVMTGLTAAFRGLAGRGOLGQVADLVPEPATVADTATPEFTER
AAGIHSVYVNGAPVWQEAFTGQHGRTVARTAA"

BASE COUNT 298 a 758 c 642 g 282 t

ORIGIN

Query Match 98.9%; Score 1739.4; DB 1; Length 1980;
Best Local Similarity 99.4%; Pred. No. 9, 1e-181;
Matches 1746; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 GAATTCACCTGATCCCGGAGAGAGATTCCTGTCCTCCATCGATCCAGCCCTTC 60

Db 168 GATCGCCTGATTCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 227

Qy 61 GACCTGTGCTCG 120

Db 228 GACCTGTGCTCG 287

Qy 121 GACCTGTGCTCG 180

Db 288 GACCTGTGCTCG 347

Qy 181 CACACCGCGGTGCG 240

Db 348 CACACCGCGGTGCG 407

Qy 241 CACGAGCACTACCTGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300

Db 408 CACGAGCACTACCTGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 467

Qy 301 ACCACGGTGTACG 360

Db 468 ACCACGGTGTACG 527

Qy 361 CCCGCCCGCTGACCTGCTGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420

Db 528 CCCGCCCGCTGACCTGCTGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 587

Qy 421 TACCTGACGCGGTTCG 480

Db 588 TACCTGACGCGGTTCG 647

Qy 481 TCAACGGCG 540

Db 648 TCAACGGCG 707

Qy 541 CGGCGCATGCGGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600

Db 708 CGGCGCATGCGGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 767

Qy 601 GGGCGCTTACACCG 660

Db 768 GGGCGCTTACACCG 827

Qy 661 CCGCTGAGCGCGCATGCG 720

[illegible]

QY	1234	ACGGCGGTTAAGAAAGATGACCCGGCTCTGACCGCGCGCGCTTCGGCCCTTGAGCCGGGCGCGG	1293
Db	1201	ACGGCGGTTAAGAAAGATGACCCGGCTCTGACCGCGCGCGCTTCGGCCCTTGAGCCGGGCGCGG	1260
QY	1294	CAGCTGCAGGCGCGGTTACTTCGCGGACCTGTTGGTTTGCACCGGCGACAGGGTGGCCGAT	1353
Db	1261	CAGCTGCAGGCGCGGTTACTTCGCGGACCTGTTGGTTTGCACCGGCGACAGGGTGGCCGAT	1320
QY	1354	ACGGCAACCTTGAACAACCTTACCGAGCGCGCGCGGCAATCCATTCCGTGACGTCAAC	1413
Db	1321	ACGGCAACCTTGAACAACCTTACCGAGCGCGCGCGGCAATCCGTGACGTCAAC	1380
QY	1414	GCGCGCCCGGTTGCGCAAGCAGGAGGTTTCAACGGCCAGCATGCGGCGCGCGGTGTCGCA	1473
Db	1381	GCGCGCCCGGTTGCGCAAGCAGGAGGTTTCAACGGCCAGCATGCGGCGCGCGGTGTCGCA	1440
QY	1474	CGCAGCGCGCGCCCTGA	1488
Db	1441	CGCAGCGCGCCCTGA	1455

```

RESULT 6  
LOCUS AF332548  
DEFINITION Alcaligenes faecalis D-aminoacylase gene, complete cds.  
ACCESSION AF332548  
VERSION AF332548.1 GI:13194735  
SOURCE Organism  
ORGANISM Alcaligenes faecalis,  
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
Alcaligenes.  
REFERENCE 1 (bases 1 to 1455)  
AUTHORS Hsu,C.S., Lai,W.L., Chang,W.W., Yang,Y.B. and Tsai,Y.C.  
TITLE Gene cloning, nucleotide sequencing of Alcaligenes faecalis DAI  
D-aminoacylase  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1455)  
AUTHORS Hsu,C.S., Lai,W.L., Chang,W.W., Yang,Y.B. and Tsai,Y.C.  
TITLE Direct Submission  
JOURNAL Submitted (28-DEC-2000) Institute of Biochemistry, National  
Yang-Ming University, 115, Li-Nong St. Sec. 2, Shih-Pai, Taipei,  
Taiwan 11221, R.O. China
```

FEATURES

source location/Qualifiers
1..1455
 `/organism=`"Alcaligenes faecalis"
 `/db_xref="taxon:"`511"
1..1455
 `/EC_number="`3.5.1.81"
 `/note="Daaf"`
 `/codon_start=`1
 `/transl_table=`11
 `/product="D-aminoacylase"`
 `/protein_id="AAK15530.1"`
 `/db_xref="GI:"`13194736"
translation="MSGPDAPPPDYIIISGTVIDGTMNAGRADNVCRKDRIIAVGDLSASSARIDIVAGKVYSPGRIDSHTHDNDITLNRDMTPRISOGVTATVGCGSLAPIAAMPRLADIIDBSGSFRFAESDIYEALAAPVAAMCVSHSTLRRAAVPDLREKTAEIQIOMQALADDALAASGAIGISTGAFPPAAHASTEEILFVCAPLTGGVYAATHRRDEGEIHVALETERIGRELDPVVVISHHKYGLNGTSGRETLLTEAMVASOVDSLDPAIPVAGSTMLOKRVLIRAGRTILTWCPRPELGSDIEELTAERKSITYDVPEPSLPAGIAYFMEDBDVDORILLAPGMGISGCLPHDRRPRLRMGITPVLYGHYSRDLIPIETAWMTKGITLAKEKLARCQGVQSEYTADVDFVPATVADSATEPLEPTERAACHISHYVGAVMWEDOSTGHAQRVNRAQA."

BASE COUNT 247 a 537 c 446 g 225 t

ORIGIN

Query Match 61.9%; Score 1088.6; DB 1; Length 1455;

Best Local Similarity 84.3%; Pred.No.6,6e+110;

Matches 1266, Conservative 0; Mismatched 229; Indels 0; Gaps 0;

```

Db 1 ATGTCAGAGCCGAGCCACCCGCTTGCATCTACTTTTCGCGCCGACACCTGTCATCGAC 60
Oy 94 GGCAGCAACACCCCGGGGGGCGCGCCGACCTTGGGCGCTGGCGGCGACCGCATCGCGCC 153
Db 61 GGCAGCAACACCCCGGGGGGCGCGCCGACCTTGGGCGCTGGCGGCGACCGCATCGCGCC 120
Oy 154 ATGCGCATCTGTCGAGCGCGCGCGACACCCGGGTGACGTGCGGGCTTGCGGCTG 213
Db 121 GTGGCGCATGAGCGCGACACACCGCGCGCGCGCATGACGTGCGGGGAAAGTGGTG 180
Oy 214 GCGCGCGGCTTTCATCTGACCTGCGACACCGACGACGACGACGACGACGACGACGAC 273
Db 181 TCTCGGGGCTTTCATCTGACCTGCGACGACGACGACGACGACGACGACGACGACGAC 240
Oy 274 ATGAGCGCCCAAGATCTCGGAGGGGCTGACACGAGTGTGACGAGGGAATGGCGCATACG 333
Db 241 ATGAGCGCCCAAGATCTCGGAGGGGCTGACACGAGTGTGACGAGGGAATGGCGCATACG 300
Oy 334 CTGGCGCGGCTTGGCGGACGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
Db 301 CTGGCGCGGCTTGGCGGACGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Oy 394 TCTTACGCTTTCGAGCGCTTTCGCGACCTGACGCGGCTTGGCGGCGCGCGCGCGCGCG 453
Db 361 TCTTACGCTTTCGAGCGCTTTCGCGACCTGACGCGGCTTGGCGGCGCGCGCGCGCGCG 420
Oy 454 GTCAAGCGCGCGCTGATGATGCGCATTCACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
Db 421 GTCAAGCGCGCGCTGATGATGCGCATTCACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Oy 514 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
Db 481 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Oy 574 GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
Db 541 GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Oy 634 ACCGAGAGATCATGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
Db 601 ACCGAGAGATCATGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Oy 694 CACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Db 661 CACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Oy 754 CCGAGCTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
Db 721 CCGAGCTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Oy 814 GCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
Db 781 GCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Oy 874 CTGAGAGCGCTATCCCTAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
Db 841 CTGAGAGCGCTATCCCTAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Oy 934 GCGGAGCGAGCAACCATCATGCTGTGCAAGCGCTTCCCGGAACTGAGCGCGCGCGCGCG 993
Db 901 GCGGAGCGAGCAACCATCATGCTGTGCAAGCGCTTCCCGGAACTGAGCGCGCGCGCGCGCG 960
Oy 994 GATGAAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Db 961 GATGAAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1033
Oy 1054 GCGGCGCGCATCTACTATCATGATGAGCAACCGCATGCGCGCATCTGCGCGCTTGGCG 1113
Db 1021 GCGGCGCGCATCTACTATCATGATGAGCAACCGCATGCGCGCATCTGCGCGCTTGGCG 1080
Oy 1114 CCGAGCATGATCGGCTGCGAGCGCTGCGCGAGAGAGGCGCGCGCATTCGCGCGCTGCG 1173

```

```

Db 1081 CCCACATGATCGGCTCCGATGGGCTGCCCAAGAGAGCGCCCGCCACCTCGCTTGG 1140
Oy 1174 GGCACCTTCCCGGGGTGCTGGGCACTATGCGCGGAGCTTGGGCTTCCCGTGGG 1233
Db 1141 GGCACCTTCCCGGGGTGCTGGGCACTATGCGCGGAGCTTGGGCTTCCCGTGGG 1200
Oy 1234 ACGGCGGTATGGAATGACCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1293
Db 1201 ACGGCGGTATGGAATGACCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
Oy 1294 CAGGTGAGCGCGCGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1353
Db 1261 CAGGTGAGCGCGCGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
Oy 1354 ACCGCGCATCTGGAACACCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413
Db 1321 TCGGCGCATCTGGAACACCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Oy 1414 GCGCGCGCGGTGCGCAAGAGACGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1473
Db 1381 GCGCGCGCGGTGCGCAAGAGACGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
Oy 1474 CCGAGCGCGCGCGCTGA 1488
Db 1441 CCGAGCGCGCGCGCTGA 1455

RESULT 7
AFAA61
LOCUS
DEFINITION
AFAA61
1497 bp DNA linear BCT 10-FEB-1999
Alcaligenes xylosoxydans DNA for N-acyl-D-amino acid
amidohydrolase, complete cds.
D45919.1 GI:1136626
VERSION
N-acyl-D-amino acid amidohydrolase.
KEYWORDS
Alcaligenes xylosoxydans (sub-species: xylosoxydans, isolate: A-6)
SOURCE
DNA.
ORGANISM
Achromobacter xylosoxydans
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
Achromobacter.
1 (sites)
REFERENCE
Wakayama, M., Ashika, T., Miyamoto, Y., Yoshikawa, T., Sonoda, Y.,
Sakai, K. and Moriyuchi, M.
Primary structure of N-acyl-D-glutamate amidohydrolase from
Alcaligenes xylosoxydans subsp. xylosoxydans A-6
J. Biochem. 118 (1), 204-209 (1995)
96015170
MEDLINE
8537313
JOURNAL
TITLE
Wakayama, M., Watanabe, E., Takenaka, Y., Miyamoto, Y., Tau, Y.,
Sakai, K. and Moriyuchi, M.
Cloning, expression, and nucleotide sequence of the
N-acyl-D-aspartate amidohydrolase gene from Alcaligenes
xylosoxydans subsp. xylosoxydans A-6
J. Ferment. Bioeng. 80, 311-317 (1995)
3 (bases 1 to 1497)
AUTHORS
Wakayama, M.
JOURNAL
Unpublished
4 (bases 1 to 1497)
REFERENCE
Wakayama, M.
Direct Submission
Submitted (13-FEB-1995) Mamoru Wakayama, Oita University, Applied
Chemistry; Danoharu, Oita, Oita 870-11, Japan
(Tel:0975-69-3311(ex.746), Fax:0975-69-7957)
FEATURES
source
1. 1497
/organism="Achromobacter xylosoxydans"
/isolate="A-6"
/sub-species="xylosoxydans"
/db_xref="taxon:85698"
1. 1497
/codon_start=1
/transl_table=11

```

/product-"N-acyl-D-amino acid amidohydrolase"
 /protein_id-"BA08350.1"
 /db_xref="GI:180111"
 /translation-"MDSRTLDPAQADFIAGATLIDGGGPARQGLAVGRIV
 ALGDAHAGVIDARGLALAPFIDSTHDDGYLAPPELPTVSGITVYVNC
 GSLAPLSRROIPOPLDLGPELPEFPAFMDRLATPAVAVIPVGTTLARA
 VMDYGRATDAERARALDLBALOAGFVSTGYTPAPAPAPDELITDQPLRG
 RAGAIYATLREADHIYAPAMEALLIGELDCRVFVSHKLAGERNHRSRTLDI
 SRAANORPCIDCHYPRATSTLRLDRALASRTLTWSKGYPEARNDSVMAELG
 LDDEAIALRLAPAGLITFLMDADVNRIFSHPLTIVSGDLPDPHPRMGSTTNV
 LRTMYREGLSLETAIHKMTGLAAQYGLERGLRQGHADLVLFDPANVTDTATE
 SAPIVSGIHAHVNRYMWDERTGERPQVLAIPGDAIPWSQSE"

BASE COUNT 235 a 592 c 467 g 203 t

Query Match 39.5%; Score 693.6; DB 1; Length 1497;
 Best Local Similarity 67.7%; Pred. No. 6.3e-67;
 Matches 988; Conservative 0; Mismatches 469; Indels 3; Gaps 1;

45 CGATTCCAGCCCTTCGACCTGCTGCTGCGGGGCGGACCCCTCATCGAGGCGAGCAAC 104
 27 CGCTCCCGGCGAAGCGATTCATATCGCCGCGCCACCTCATCGAGGCGGCGCG 86
 105 CCGGCGGCGCGCGCGCGACCTGCGCGCGCGCGACCGCATCGCCGCGCATCT 164
 87 CCGCTT 146
 165 GTCGAGCTT 224
 147 CGCCCATCGCTT 206
 225 CATCGACTCGCGACCGCGAGCGACACTACCTGCTGAGCGCTGCGCGCATAGCGCCAA 284
 207 CATCACTCGCGACCGCGAGCGAGCGGCTACTGCTGCGCGCATCCCGAGTCTGCCAA 266
 285 GATCTCGAGGCT 344
 267 GGTTCGAGAGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 326
 345 GCGCGACCGCTT 404
 327 GTCGCGGCGCGAGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 386
 405 CGAGCGCTCGCTT 464
 387 CGCCACCTTCGCTT 446
 465 CTGTATGCTGCTT 524
 447 CCGCTGCTGCTT 506
 525 CACCGAGGAGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 584
 507 CACCGAGGCTT 566
 585 CATCGGCTT 644
 567 CTTGCGGCTT 626
 645 CATCGAGGCTT 701
 627 CATCGAGGCTT 686
 702 CGAGGAGGCTT 761
 687 CGAGGAGGCTT 746
 762 GCGAGGCTT 821
 747 GCGAGGCTT 806
 822 GCGAGGCTT 881

Db 807 CCGGAAAGCGTGGAGCATGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 866
 882 GATTCCTTCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 941
 867 CCACCCCTTCAGGCTG 926
 942 CACCATCATCATCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1001
 927 CACCCGATCATCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 986
 1002 CCGGCT 1061
 987 CATGCT 1046
 1062 CATCTACTCATGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1121
 1047 CATCTACTCATGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1106
 1122 GATCGGCTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 1181
 1107 GGTGCGGCTT 1166
 1182 CCGGCGGCTT 1241
 1167 CACCAACCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 1226
 1242 ATGGAATGAGCTT 1301
 1227 CCACGAATGAGCTT 1286
 1302 GCGCGGCTACTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1361
 1287 GCAAGGCTATCATCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1346
 1362 CTTGGAACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 1421
 1347 TTTTTCAGCTTCCATTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1406
 1422 GGTCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 1481
 1407 GGTCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 1466
 1482 CCGCTGAGCTT 1501
 1467 TGCAATACCTTGGAGTCAAC 1486

RESULT 8
 AL646083
 LOCUS 208050 bp DNA linear BCT 07-DEC-2001
 DEFINITION Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
 accession AL646083
 version AL646083.1 GI:17431588
 keywords Ralstonia solanacearum.
 source Ralstonia solanacearum.
 organism Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 Ralstonia.
 1 (bases 1 to 208050)
 Salenoudet,M., Genin,S., Attiguenave,F., Gouzy,J., Mangenot,S.,
 Arlat,M., Billault,A., Brotlier,P., Camus,J.C., Cattolico,L.,
 Chandelier,M., Choisme,N., Claudel-Renard,C., Cunne,S., Demange,N.,
 Gaspin,C., Layle,M., Molan,A., Robert,C., Saurin,W., Schlex,T.,
 Siglier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
 Weissenbach,J. and Boucher,C.A.
 Genome sequence of the plant pathogen Ralstonia solanacearum
 Unpublished
 2 (bases 1 to 208050)
 Boucher,C.A.
 Direct Submission
 Submitted (05-DEC-2001) Genoscope and CNRS UMR 8030, 2 rue Gaston
 Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT
Molculaire des Interactions Plantes-Microorganismes INRA-CNS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS,
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, 31326 Castanet-Tolosan Cedex
Christian.Bouchet@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES

Location/Qualifiers
1..208050

source

/organism="Ralstonia solanacearum"
/strain="GM11000"
/db_xref="taxon:305"
/plasmid="megaplasmid"

gene

135..947
/gene="Rsp1115"
/note="RS05505"

CDS

135..947
/gene="Rsp1115"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"

gene

/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"

CDS

/protein_id="CAD18266.1"
/db_xref="GI:17431589"

gene

/translacion="MADSVQVQVADYDGLVLSAGRSQSFERRHPLDPHGLVATAG
VORRFRGKTVLPGPCISILMPGEIHIDGVAEGGATLTFTPLSOLALANAGADISG
TDHEPELATLLEDDPALAGHLRLHDMHSGASTLOVQSEMLALFLILKQADILP
EHTKGLSPAMQOVRYDRCFAHGERETDLDASLGLGFRQFKQPKQVGTPIPAW
LVRLRQAGSLSSRSKALAEVAQEVGDFYDQSHFNRAFRQAGVAPSHYS"

CDS

1162..2259
/gene="Rsp1116"
/note="RS05504"
1162..2259
/gene="Rsp1116"
/EC_number="4.1.2.5"
/function="small molecule metabolism; degradation; amino
acids"

gene

/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"

CDS

/evidence-not_experimental
/transl_table=11
/product="PUTATIVE LOW SPECIFICITY L-THREONINE ALDOOLASE
PROTEIN"

gene

/protein_id="CAD18267.1"
/db_xref="GI:17431590"

misc_feature

2257..5019
/note="ACUR or RS05503
Predicted by Framed"
/evidence-not_experimental
complement(2462..3139)
/gene="Rsp1117"
/note="RS05502"
complement(2462..3139)

gene

/note="ACUR or RS05503
Predicted by Framed"
/evidence-not_experimental
complement(2462..3139)
/gene="Rsp1117"
/note="RS05502"
complement(2462..3139)

CDS

complement(2462..3139)

CDS

/gene="Rsp1117"
/function="miscellaneous; unknown"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"

gene

/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PUTATIVE TRANSMEMBRANE PROTEIN"

CDS

/protein_id="CAD18268.1"
/db_xref="GI:17431591"

gene

/translacion="MNDASLHETIISCFLEHOPRPTREIAERPNDEQAVYRALRAL
ADNHGYLHPHSDVWIAHPESAPTCYKSGRRKMGCMACALGLVLAGSATI
ETRGADIDHDIYIRIENGRLDIDYVHFPIPKHADVITYCSVOLFRDRAOYDE
WASRGIPKGDVRIEYWNFAVEMVARRHADADWTMSVREAIELFARHLLSGPVWTI
AAEAR"

CDS

complement(3188..3595)
/gene="Rsp1118"
/note="RS05501"
complement(3188..3595)
/gene="Rsp1118"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"

gene

/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"

CDS

/protein_id="CAD18270.1"
/db_xref="GI:17431593"

gene

/translacion="MGWSEGEYCPARTAVYGDKRTVITLLEFLGYKREPEFOA
OTGMSHLSRRLKRLPDGIVVRLHLSDRPREYRLTSGDLPYLLSLKSGCK
WGFRAKATAPALVITTHRCGHEGLKIVCPACDPEFQKATVTLGASFAEROARD
GA"

CDS

4228..4725
/gene="Rsp1120"
/note="RS05499"
4228..4725
/gene="Rsp1120"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"

CDS

/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"

gene
/protein_id="CAD18271.1"
/db_xref="GI:17431594"
/translation="MPATRRPREGAPRTLEDKFKVDTLETSGEAVTRDMILSPAEQY
DPOPMHDLARADTVFGLVSGWGLIADRLDRLGIPVAGELRDLREHA
PMRGDVLRAAEVIAIRPSKSRPDRGMARVTTTNSGVTLYTQKMSLVPRPDR
QKVAE"
/gene="RSP1121"
/complement(5017, .5610)
/note="RSP1121"
/function="miscellaneous: not classified regulator"
/note="Product confidence: putative
Gene name confidence: hypothetical
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not-experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD18272.1"
/db_xref="GI:17431595"
/translation="MAEKITRRPRADAENRRLDPAVTAFAEKVSASIDIAE
AGVIGILYRHPPTDRLIDELIYRDESSRLAEAROSEOPPLEAVRQMLLFVGL
ANKOIRADVINCMTDSEKICISGELYITLQILIRAKOSGATGLAVEPDLISAV
AGVASFGEIDMNAEGRVLEVMVAGLRVAATGGRK"
5822, .6910
/gene="RSP1122"
/note="RSP1122"
/note="RSP1122"
/note="RSP1122"

Query Match 30.44; Score 533.6; DB: 1; Length 208050;
Best Local Similarity 61.38; Pred. No. 2.9e-50;
Matches 876; Conservative 0; Mismatches 549; Indels 3; Gaps 1;

gene
57 CTTCGACCTGCTGCTCGGGGCGGACCTCTATCGACGCGAGCAACCCCGGGGCGG 116
Db 95887 CATCCGAGGAGTCCGATCGATCGACGCGAGCGGCGCCAGCTCGTTGCA 95946
Qy 117 CGCGGACCTGGGGTGGCGGCGGACCGCATCGCCGCGCATGTGTCGAGCGCG 176
Db 95947 CGTCGCGATCGCGATGCGCGATCGCGCGCGCATCGCCGCGCATCGCGCG 96006
Qy 177 CGCGGACCTGGGGTGGCGGCGGACCTGTGCGGCGCGCGCGCGCTTCATCGAC 236
Db 96007 GGGCGAGGAGAGTGCAGCGCGCGCGCGCGCGCGCGCGCTTCATCGAC 96066
Qy 237 CACCGACGAGACACTTACTGCTCAGGCGCTCGACATGACCGCCCAAGATCGCAGG 296
Db 96067 CACCGACGAGACACTTACTGCTCAGGCGCGCGCGCGCGCGCGCGCTTC 96126
Qy 297 CGTCGACGAGTGGTGCAGCGGCAATGCGCGCATCGCGCGCGCGCGCGCTTC 353
Db 96127 CATCACCACGATGATGCGGCAATGCGCGCATCGCGCGCGCGCGCGCGCT 96186
Qy 354 CAACCGCT 413
Db 96187 CGAGCT 96246
Qy 414 CGCGGACTTACTGAGCGCTTGGCGGCGAGCGCGCGCGCGCGCGCGCGCGCT 473
Db 96247 CGCGGACTTACTGAGCGCTTGGCGGCGAGCGCGCGCGCGCGCGCGCGCT 96306
Qy 474 GGGCGATTCAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 533
Db 96307 CGGCGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 96366
Qy 534 GGAATTCGCT 593
Db 96367 CGAAGTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 96426
Qy 594 TTCGACGCT 653

Db 96427 GTCACCGGACACGCGCTATGCCAACGCCCATGCGCGCGCGCGCGCGAGATGAGCT 96486
Qy 654 GTCCGCGCGCGCTGAGCGCGCATGGCGCATCTACGCCACCCACATGCGCGAGAGGGA 713
Db 96487 GCGCGAGCT 96546
Qy 714 GCACATCGTGGCT 773
Db 96547 GCGCATCTGGAGCGGATGAGAGAGAGCTTACCGGCGCGCGCGCGCGCGCGCT 96606
Qy 774 GGTGATCTGCGACACCAAGATCATGGCGCGCGCGCGCGCGCGCGCGCGCGCT 833
Db 96607 CGTATCTGCGACCGTGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 96666
Qy 834 GCGGTGATCGAGCT 893
Db 96667 GGAATTCGAGCT 96726
Qy 894 GCT 953
Db 96727 CGCGACGCTGCGACCGCTGACCTGAGCTGAAGAGTGAAGAGATTCGATCGTATCAC 96786
Qy 954 CTGCTGCAAGCGCTTCCCGGAACTGACCGCGCGCGCGCGCGCGCGCGCGCGCT 1013
Db 96787 CTGCTGCGACCGCGCATCGGAGATGGCGCGCGCGCGCGCGCGCGCGCGCT 96846
Qy 1014 CGGCAATTCAGTACAGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1073
Db 96847 GGGGACGACCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 96906
Qy 1074 GATGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1133
Db 96907 CATGCT 96966
Qy 1134 CGGCTCTCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1193
Db 96967 CGGCTCTCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 97026
Qy 1194 GGGGCACTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1253
Db 97027 CGGCGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 97086
Qy 1254 CGGCTCTGCT 1313
Db 97087 CGGAGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 97146
Qy 1314 CGCGGACTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1373
Db 97147 GCGCGACTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 97206
Qy 1374 TACGAGCT 1433
Db 97207 CAGAGCT 97266
Qy 1434 GCAAGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1481
Db 97267 CGGTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 97314

RESULT 9
AFNA6
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

1467 bp DNA linear BCT 10-FEB-1999
Alcaligenes xylosoxydans DNA for N-acyl-D-amino acids
amidohydrolase, complete cds.
D50061.1 GI:769683
N-acyl-D-amino acids amidohydrolase.
Alcaligenes xylosoxydans (strain A-6) DNA, clone pAD.
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
Achromobacter.
Wakayama, M.

TITLE Studies on structure and function of bacterial N-acetyl-D-amino acids
amidohydrolase
JOURNAL Thesis (1995) _FACULTY_/Applied Chemistry/_MAJOR_, Oita University
REFERENCE 2 (bases 1 to 1467)
AUTHORS Wakayama, M.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1995) Mamoru Wakayama, Oita University, Applied
Chemistry, Danonoharu, Oita, Oita 870-11, Japan
(Tel:0975-69-3311(ex.746), Fax:0975-69-7957)
Location/Qualifiers
source 1. 1467
/organism="Achromobacter xylosoxidans"
/strain="A-6"
/db_xref="taxon:85698"
1. 1467
/codon_start=1
/transl_table=1
/product="N-acetyl-D-amino acids amidohydrolase"
/protein_id="BA08778.1"
/db_xref="GI:1707291"
/translation="MOEKLDVIEGWVIDLGGPRRRADYVIGERIALTSDLSAAP
ADRLDAGGRIVAFIDTGHDDLMFEKELMKTSOGITSVYVCGISGAPL
PGNTAALALIGDPLFADAMATRGALAEADBPMTNVALNHNALRIAMADPAOS
AKBORAMERMLADALEAGAVSTGLAYOPGVABQELDGLANVAARGLHTSHR
NEBPAYEADEVTAAGRTGCAIVLSHKCMAMPNGKSAATLANIDRARAAGVDA
LDITPFGSSSTILPERADQIDIRITWSPPHPCGGSLAEIARNGCDVTAARL
CPAGATFAMDENENVRIFQHECCMGVSDPNDPNAHVPDADYVCDRAITWDAFTLASGI
LLEAAVAKMTALPARVFGGLADRGRLAVGAMADYVADADYVCDRAITWDAFTLASGI
EHVNVNCAVFPQAPSPHRRGRITRRDASTAGAEFSR"
BASE COUNT 213 a 530 c 530 g 194 t
ORIGIN

Query Match 25.9%; Score 455.8; DB 1; Length 1467;
Best Local Similarity 59.3%; Pred. No. 4.8e-41;
Matches 813; Conservative 0; Mismatches 552; Indels 6; Gaps 2;

59 TCGACCTGCTGCTCGCGGGCGGACCCCTCATCGACGCGCAACACCCCGGGCGCG 118
14 TGGATCTGTGATCGAAGCGCGCTGGGTGATCGAGGCGCTGGGCGCGCGCGCG 73
119 CGACCTGGCGCGCGCGACGACGATCGCGCATCGGGATCTGTGCGAGCGCGCG 178
74 CGAGCTGGGATCGGTGGCGACGCGATCGCGCATCGCGACCTGTGGCGCGCGCG 133
179 CGCACACCGCGGCTGAGCTGTGCGGCTGTGCGCGCGCGCTTCACTGACGACA 238
134 CGGACCGCGCGCTGATGCGCGCGCGAGATCGTGGCGCGCGCTTATGATACCATG 193
239 CGCACGACGACACTACTGCTCAGCGGTGCGACATGACGCCCAAGATCTGCGAGGCG 298
194 GCCACGATGACCTGATGCTGGAACAAACCGGCGCTGGAATGGAACGACGAGGCA 253
299 TCACACGCTGTGTCACGGGCAATGGGCAATGACCTGGGCGCGCGCTGGCGACCGCA 358
254 TCAGCTGGTGTGTGGCGCACTGCGGCACTACGCGCGCGCGCGCGCGCGCGCGCA 313
359 CGCGCGCGCGCTGACCTCTGAGACGAGAGCGGCGCTTACCGTTTGAAGCGTTGGCG 418
314 ATACCGCGCGCGCGCGCTGTGGCGACCTGCGCGCTGTTCGCGCAATGCGCA 370
419 ACTACCTGAGACGCTTGGCGGCGACGCGCGCGCGCTCAAGCGCGCTGTATGGTGGCG 478
371 TGTACTTGGCGGCTGAGAGCGCGCGCGCGATGATCAACGAGCGCGCGCTGTGGCG 430
479 ATTCAACGCTGGCGCGCGCGCGCTGATGCGGACCTTGAAGCGCGCGCGCGCGCAAGAAA 538
431 ACGCCAAGCTGGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 490
539 TCGCGCGCATGCGGAGACTGCGCGAGAACGCAATGGCGAGCGCGCGCGCGCGCGCGCG 538
491 AGCGGGGATGAGCGGATCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 550
599 CGGCGCGCTTCTACCG 658

551 CGGGCTGGCTTACAGAGCGCGCGCGCGCTGCGAGCAAGCGCACTGAGCGCGTGGCG 610
659 GCGCGCTGAGCGCGCATGGCGCGCATCTACGCCACCCACATGCGCGAGCGGACGACA 718
611 GCGTGGCG 670
719 TGGTGGCG 778
671 TCGAGCG 730
779 TCTCGCACACCAAGGATGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838
731 TGTTCGCACCAAAATGATGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790
839 TGTATGA---GCGCGCGCATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
791 ACATGACCG 850
896 CGGCGTCCACCATGCTCAAGAGAGACCGCGTCTGCTGCGCGCGCGCGCGCGCGCGCG 955
851 GCGATTCACCATCTGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910
956 GGTGCAAGCCCTTCCCGCAACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
911 GGTGCAAGCCCGATCCCGAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
1016 GCAATTCAGATGACGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075
971 GCTGCGAGCGGTCACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1030
1076 TGGACGACCG 1135
1031 TGGACGACGAGGTCG 1090
1136 GCTGCG 1195
1091 GCTGCG 1150
1196 GGCATATGCG 1255
1151 GCGCGTACGTGCG 1210
1256 GCGTGAAGCG 1315
1211 GCGTGGCG 1270
1316 CGGACCTGTGTGTTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1375
1271 CGGACGTGTGTGTGTTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1330
1376 CGGAGCG 1426
1331 CGTGGCG 1381

RESULT 10
LMEFLCHR34_10/c
WPCOMMENT
Sequence split into 18 fragments LOCUS LMEFLCHR34 Accession AL499623
Fragment Name Begin End
LMEFLCHR34_00 1 110000
LMEFLCHR34_01 100001 210000
LMEFLCHR34_02 200001 310000
LMEFLCHR34_03 300001 410000
LMEFLCHR34_04 400001 510000
LMEFLCHR34_05 500001 610000
LMEFLCHR34_06 600001 710000
LMEFLCHR34_07 700001 810000
LMEFLCHR34_08 800001 910000
LMEFLCHR34_09 900001 1010000
LMEFLCHR34_10 1000001 1110000
LMEFLCHR34_11 1100001 1210000

small overlap between neighbouring submissions. Cosmid 2K36 overlaps cosmid SCK15 and cosmid SCK13.

```

FEATURES
    source
        1..22251
            /organism="Streptomyces coelicolor A3(2)"
            /strain="A3(2)"
            /db_xref="taxon:100226"
            /clone="cosmid 2K36"
            3..112
                /note="nominal overlap with Streptomyces coelicolor cosmid SCK13"
                complement(160..864)
            /gene="2SCK36.01c"
            /note="SC04978"
            complement(160..864)
            /gene="2SCK36.01c"
            /note="2SCK36.01c, conserved hypothetical membrane protein, len: 234 aa; similar to SW:YAB8_MTCU26.0 kDa protein RV1085c or MV017.38c, 242 aa; fasta scores: opt: 563 Z-score: 653.6 bits: 128.3 E(): 8.6e-29; hydrophobic membrane spanning regions"
            /codon_start=1
            /transl_table=11
            /product="putative integral membrane protein"
            /protein_id="CAC40591.1"
            /db_xref="GI:14285272"
            /db_xref="SPTREMBL:093JL6"
            /translation="MTASVPDAPRDLPAGRGPTVLSLPHYKPKRLKGLHLMGMPAYLVAGVLTALAEITSEARICGVYVITACILGVSLHYRGMSPMDQVLRDLANIWAAYVTEPBRAGGVAVLVITVGGILYSAGVLYGKRPNSPRMGFEVHSPT1141..1145
            1151..2980
            /gene="2SCK36.02"
            /note="SC04979"
            1151..2980
            /gene="2SCK36.02"
            /note="2SCK36.02, probable phosphoenolpyruvate carboxykinase, len: 609 aa; similar to SW:PECK_CHLII(EMBL:S56812) Chlorobium limicola phosphoenolpyruvate carboxykinase [GFP] (EC 4.1.1.32) PCNA, 646 aa; fasta scores: opt: 2400 Z-score: 2689.0 bits: 507.7 E(): 3.6e-142, 58.3618 identity in 598 aa overlap. Contains Pfam match to entry PF00821 PEPC, Phosphoenolpyruvate carboxykinase and match to Prosite entry PS00505 /codon_start=1
            /transl_table=11
            /product="putative phosphoenolpyruvate carboxykinase"
            /protein_id="CAC40592.1"
            /db_xref="GI:14285273"
            /db_xref="SPTREMBL:093JL5"
            /translation="MARDAIAPPPVPRNHQELISWNEIALETOPDAVWVDCSEAEYERLCELYEKGTPRKIDPIKRNSTYASDPITVAREDTFICSAKEEAPGNTNMDKPMTRMAGIDVDELDEGEFVAVASVGALEPGQADVPNCNRSKYSHPFEDIEINSGSYTGAGNALGKRCYALIRIVARDEGMALAHMLYKILPTPGARVYAAAPSA NAMKTLGNSVETVALTDGDDVWMEGTEBETPAHLDWKGDWTPESGTPAANARFTTAPAOCPITAPEDWPRGPVPSILILGGRATVAPLVYTESFDNNGVFLGANVASE DAGEFVMPGFGENGKLVMIIVERLEGADGVETPIGVLPTEKESLDTDLADLADLER ILSDVKEVYREELALVPEHLNFTFGDHPFAELMDYRALVRLG"
            1205..2974
            /gene="2SCK36.02"
            /note="Pfam match to entry PF00821 PEPC, Phosphoenolpyruvate carboxykinase, score 1344.70, E-value 1967..1993
            /gene="2SCK36.02"
            misc_feature
                misc_feature
                    misc_feature

```

```

gene
    /note="PS00505 Phosphoenolpyruvate carboxykinase (GTP) signature"
    complement(3263..4213)
    /gene="2SCK36.03c"
    /note="SC04980"
    complement(3263..4213)
    /gene="2SCK36.03c"
    /note="2SCK36.03c, unknown, len: 316 aa"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein 2SCK36.03c"
    /protein_id="CAC40593.1"
    /db_xref="GI:14285274"
    /db_xref="SPTREMBL:093JL4"
    /translation="MSGDGLEGLAAVHRLTAQRMGPVYCGTAGHGTGVCVTPAGAGLPGLIEAIGSRGEPRLVATGTTTPAAGSAGLPLRTGEPRIYKMAWAGDRKIGCTVRADDDADGAVRPVYLVAREADPAAGASAPRTWDGIVATGRTAARAVMDPAV EAKLGTPLDGYKRLVEIFGAGAGVYQLOLPDADHSDTDLVRETEEMGEMARTG SLMERYPYRPAAGGLQAWASVGEADGFYRLVGDGPDVPLANEVDPSWEREDGP TGEFVRMLTEPDHPSTARNVDVHTDYGPAADEGRV"
    complement(4210..3610)
    /gene="2SCK36.04c"
    /note="SC04981"
    complement(4210..3610)
    /gene="2SCK36.04c"
    /note="2SCK36.04c, possible bifunctional protein, len: 466 aa; N-terminal region presents low similarity to N-terminal region of TR:O14362 (EMBL:297992) Schizosaccharomyces pombe putative glucan synthase-1 (EC 2.4.1.34) SPBC30D10.17c, 504 aa; fasta scores: opt: 176 Z-score: 188.3 bits: 44.3 E(): 0.0071; 24.931% identity in 361 aa overlap and C-terminal region to C-terminal region of TR:AK22951 (EMBL:AE005774) Caulobacter crescentus oxidoreductase, GFO/IDH/MOA family CC0967, 307 aa; fasta scores: opt: 146 Z-score: 160.2 bits: 38.4 E(): 0.26; 26.590% identity in 173 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="putative bifunctional protein"
    /protein_id="CAC40594.1"
    /db_xref="GI:14285275"
    /db_xref="SPTREMBL:093JL3"
    /translation="MSISIFATWPELRLRYASHPERLAGPGHVGQITGLGWSVPVRPHAPGRASLVEDMODEFTVEAQAALRAHGRQSVSEFVETVMDALTAHVDG GRAVEHGVSPAGTIVLYEGANPEWRKRLPEVNPALPAPSDALLEETLIERLD LQGLYADARSRCMLQGLVLPDELKALRYVARRAGMDKGDVYAGQRYDAVCCELP PPGGHTGOVVMIGHEDSIGAGLLASLDMDVNGRAEHRGRDNDPVPVARNLVG DVSAAVARELEVLYVGGREGEPRSLAPAGLPRLTLACGALADPLEIGLTRLE YRLGPEERVLVDGAVPRLSLAAHIEVGEHHPRIITALLANELSLMDRPLISRTV LEGDDTDRTAGGAR"
    complement(5618..5621)
    5786..5792
    5801..5983
    /gene="2SCK36.05"
    /note="SC04982"
    5801..5983
    /gene="2SCK36.05"
    /note="2SCK36.05, putative membrane protein, len: 60 aa. Contains possible hydrophobic membrane spanning region"
    /codon_start=1
    /transl_table=11
    /product="putative membrane protein"
    /protein_id="CAC40595.1"

Query Match
    9.8% Score 171.8; DB 1; Length 22251;
    Best Local Similarity 49.5% Pred. No. 1.5e-10;
    Matches 813; Conservative 0; Mismatches 647; Indels 183; Gaps 7;

OY 32 CCATGCCCATCCGATCCAGCCCTTCGACTGCTGCGCGCGGACCCCTCATCG 91
DB 10533 CCGGAGCCAGAGAGATCAGGCATGGAAGAGCTGTCATCAGGACCGGACGCTCTCG 10474

```

REFERENCE	TITLE	JOURNAL	FEATURES
12	AF453501/c	LOCUS	82746 bp DNA linear BCT 18-JUN-2002
12	AF453501	DEFINITION	Actinosynnema pretiosum subsp. auranticum maytansinoid antitumor agent anasmitocin biosynthetic gene cluster I, partial sequence.
12	AF453501	ACCESSION	AF453501
12	AF453501	VERSION	AF453501.1 GI:21449342
12	AF453501	KEYWORDS	Actinosynnema pretiosum subsp. auranticum.
12	AF453501	SOURCE	Actinosynnema pretiosum subsp. auranticum.
12	AF453501	ORGANISM	Actinosynnema pretiosum subsp. auranticum.
12	AF453501	ORGANISM	Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae; Actinomycetales; Pseudonocardiales; Actinosynnemataceae; Actinosynnema.
12	AF453501	REFERENCE	1 (bases 1 to 82746)
12	AF453501	REFERENCE	Yu, T.-W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.O., Xu, J., Moss, S.J., Leistner, E. and Floss, H.G.
12	AF453501	REFERENCE	The biosynthetic gene cluster of the maytansinoid antitumor agent anasmitocin from Actinosynnemapretiosum
12	AF453501	REFERENCE	Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7968-7973 (2002)
12	AF453501	REFERENCE	2 (bases 1 to 82746)
12	AF453501	REFERENCE	Yu, T.-W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.O., Xu, J., Moss, S.J., Leistner, E. and Floss, H.G.
12	AF453501	REFERENCE	Direct Submission
12	AF453501	REFERENCE	Submitted (28-NOV-2001) Department of Chemistry, University of Washington, Box 351700, Seattle, WA 98195-1700, USA
12	AF453501	REFERENCE	Location/Qualifiers
12	AF453501	REFERENCE	1. 82746
12	AF453501	REFERENCE	/organism="Actinosynnema pretiosum subsp. auranticum"
12	AF453501	REFERENCE	/strain="ATCC 31565"
12	AF453501	REFERENCE	/sub_species="auranticum"
12	AF453501	REFERENCE	/db_xref="ATCC:31565"
12	AF453501	REFERENCE	/db_xref="taxon:42198"
12	AF453501	REFERENCE	/note="cluster I of the maytansinoid antitumor agent anasmitocin biosynthetic gene cluster"

```
gene
<1.1678
/gene="asm37"
CDS
<1.1678
/gene="asm37"
/codon_start=2
/transl_table=11
/product="unknown"
/protein_id="AAM54115.1"
/db_xref="GI:21449383"
/translation="1TPGLVVALAANGVLAAPASAAITSPVTSTATVTEFVSRGT
PAFSRYVDTDSAAANGVNOGSGTGDHLENGSLRHSGGGAWMTQVETVPTSAAGV
ARKYTERADLGEASPNVDVYFOVEAPVETSAKTHHTYAGAGSVAYQPSAVNP
PENGILHTGDCDKADLSADALRRYRTGQITLACVEYLAHRSSPIAQTALDLOR
QFDDAAVAGCVKLYLRFAYTASESGEDAPKERYLAIDOLAPLIRANADYTAVOGCV
GAMGEVYTKNFGNAGVYSATDOANKAVYDKLTAIPADRAVOLTKRKRAALGGG
STRIGHHNCDFLSDPTDEGTEDPATREYLYLEAERTPLMGECATVAMPRTDCTAT
AELSRHMTYLVTDYQODVLRWMSNOGCLAEYTAQTLGRTLRDSVFPPTTRGALP
VRVNLNDGYAAAVQRTVNLRLNTTIGALTRLPLTADPRTMLPGARAIADQADIPP
SLATQIALLELADPALPGREYSIQYANTGLMDARTGLNDLKQITIVS"
1825.2142
/gene="asm36"
1825.2142
/gene="asm36"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54114.1"
/db_xref="GI:21449382"
/translation="MTASAVICARAGCGAPRPGATARKFFCSTACKQAAVRLRRA
GSASHEHDEQLREKLALVVARLTLTEATPWSAQATALVGLARVEVLAADAGV
SGG"
complement(2129.3307)
/gene="asm35"
complement(2129.3307)
/gene="asm35"
/codon_start=1
/transl_table=11
/product="bicyclomycin resistant protein"
/protein_id="AAM54113.1"
/db_xref="GI:21449381"
/translation="MAVGKADSAVAGNLPALLIYALNGTIIIDMSVSPEDTAREIG
VGGPAVOLTVTLNFGICGALGCHGSPDRGRPVILGSAWVLAALCALAPLTIG
LLTARFVQIGASAAVIVITIGDYSRBPALRMATGLVNAASALMTACAPLGGAI
NRVAGNGNYLACATATLSILVFLPEKTRATITARSVLAADYRLLSRAVL
AASITPILRAIYVTAASPELTAAFGVGFALHLLVAVASFAVPAFAITGIIP
LLGGPAGAVRWGAGVAVVGLGFLLDGCVAVTGCVFAGVVCYVPVPRSMKV
FPBLGAASSATMGLRALVTLTLTALAGALSTGCPAPAAVMAVAVLLLSLSR"
complement(3450.4064)
/gene="asm34"
complement(3450.4064)
/gene="asm34"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAM54112.1"
/db_xref="GI:21449380"
/translation="MAKRNPERRAALLDAEVLAGEAGRLTFRAVDORGAVPACTA
SNTFAGAEILITGAGERVYORLDOAPLIGETPGPDRSRVTEIMHALVDVSAFPG
FLALBELRLAARNPELREVLTAARVADLRFNENHHAASGLPGDCTVALLMLANML
ILERLTLPDLTTHORRDLYTALVVDLLAAGPPEASPFDESCK"
4151.4723
/gene="asm33"
4151.4723
/gene="asm33"
/codon_start=1
/transl_table=11
/product="dihydrofolate reductase"
/protein_id="AAM54111.1"
/db_xref="GI:21449379"
/translation="MRKLVYFIASIDGYIAPGECHDFPTAEPVVEHLRAEYPTT
PTHVRPLVGMPLDTPNRRYDTVLWGRSGVPLLDGVSFVAHLRQYVYSRLPADP
AVALVADPVGIVRELKREDGLDVMLCGGADLAGLLPIDEIVKTFIVVAGGTPPL
```

```
gene
FRAGFAPRAFRPVDVRAFDHGGVTTTYPAA"
complement(4854.5057)
/gene="asm32"
complement(4854.5057)
/gene="asm32"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54110.1"
/db_xref="GI:21449378"
/translation="WTEQNAKIPARSSGRSNAARCAVMPRPQPSATRAPPRS
SANDSRRSRGRDPRISLNSA"
5415.6461
/gene="asm31"
5415.6461
/gene="asm31"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor"
/protein_id="AAM54109.1"
/db_xref="GI:21449377"
/translation="MWIGYSADGRRRPVPMWGVTEAFWTEDDTLYALTTRDAFA
AMADSTRAYLKAIGRRLITADDEVLEARVEGLTAEHLIDLPAPARSRRVEAR
RYVDDGGRADHLEANLRLVVSIAKRTYKGMALLDVIOGNGLVRAVEKEDYARG
EKSTIVATWMLRQAITRALDQSTRTILPVHMEVYTKLARIEDLHNRGRTAE
VAVELDVTPEKVELROHARTPVSLDITIAEDTSIGDLIDENAVAPLEAVSALL
SADRLALETLEPREAGVVRNRGFPDGVLTHTIDEVGRVYGVTRERIROTESRTMALT
OHPARSRLGEGYLD"
complement(6501.9518)
/gene="asm30"
complement(6501.9518)
/gene="asm30"
/codon_start=1
/transl_table=11
/product="cytochrome P450"
/protein_id="AAM54108.1"
/db_xref="GI:21449376"
/translation="WATGTRIPGPRPLVGNLIDLVTLSDDTDDVDLDRCHREBGG
ITVALTRGQOVQVASSHELVAARCSDPSWKAHPALAEQVRDAGGLFARDEPWW
GKHRLIMPFEQYDRAHDEHPAMLIDAEOMLVMRREPGDPRIDVADNDRLDITLA
LCFEGARFNFDYDRAHDEFPVDAVRSIVGEARELPGVPTLVGRNDRYRDITMT
NRIADQIVARALPAGEPDDLEMLRLCADPVEGRSARVRYQLATFLTAGHET
TSGLSFVARRILAHPEVLRKADANDVULGDVPAFEDLARLDYLGQVIRETLRHP
TAPAFIADPERBELGCHAGCEPVLMPLTHRDPVNRDQVDPVDFPFAERDE
IPCAAMPFGHGKACTIGREFALQEPATLVLAQLQFDLADPDHRLTIKQTLTKP
DSLVAARPDGPRGATAVEITVPQVATHRGTPLVHFGSNGSGEGALRTTAG
DGAARGMATSVAPLDDAVRALPASGVVYVSSYNGAPPDNAHFVRMTTODGPDLSG
VDLYLVCGNLDSATYORVPTLIDEMAAAGARRLRBGATDARADPFGDMEKRTES
LMPILSAGGVGEIGEPFRFVYESAAGDGLDASVYLENRELVRGPDAGSKRILE
LRLPDGSTRGTGYLSVLPQNHEDLVRRVARLGTAEVRYVESAPGLVPGVAL
RVDDELTRCVDSLAPAGAGVVARLAERCCPPERELATATGATLLELLEPSCAVD
LALAEILLPAPRTRLYSISSAEQRAEVALVSYGVTSGLSRPGRDRAVAGIAS
PESFERPADNTVPVYLIAAGTIGAFEGFLNARALAGEPGLALLPFCRCRPEIDDL
VAEEFALDMLDEVRAYSRHPDGEVRHQHLMRDRDRELYNAGARYLCSGNATR
VGAPEEVLGRIQPGAGWMLRAGGRATYDVF"
9664.10245
/gene="asm29"
9664.10245
/gene="asm29"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAM54107.1"
/db_xref="GI:21449375"
/translation="MSAAVREQLVAALRVKRDGIAAATTAICABADMPHGAFHY
GERSKOEYLTALLTGLNTSLDAAAPASLPDPAVNDRTMLAANRTAFOADPRELV
LGEIVTLAOREBELRELPAHOMREVVAAVAPDLALAEAGPDISVDALALIELVSV
LDGVASMFSSRDGLAASLDQFVAFALNR"
10393.10857
/gene="asm28"
10393.10857
/gene="asm28"
/codon_start=1
```


VERSION ALL38978.2 GI:20520797
KEYWORDS alpha-amylase; g1991, 1/4-alpha-glucan branching enzyme; g19P, glycogen phosphorylase; integral membrane transport protein; pfk2, 6-phosphofructokinase; pta, phosphate acetyltransferase; secreted nucleosidase; sodium:dicarboxylate symporter; transcriptional regulator; trkA2, thiodoxin; two-component sensor kinase/sensor kinase.

SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
REFERENCE Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.
AUTHORS 1. (bases 1 to 28751)
Redenbach, M., Kiese, H.M., Denaplatte, D., Elchner, A., Cullum, J., Kinashi, H., and Hopwood, D.A.
A set of ordered cosmid and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
JOURNAL 97000351
MEDLINE 8843436
REFERENCE 2. (bases 1 to 28751)
Olliver, K. and Harris, D.
AUTHORS Unpublished
JOURNAL 3. (bases 1 to 28751)
Bentley, S.D., Parkhill, J., Barrett, B.G. and Rands, M.A.
REFERENCE Direct Submission
JOURNAL Submitted (03-FEB-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SQ E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
On May 9, 2002 this sequence version replaced gi:6941966.
COMMENT Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Wellcome Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/projects/S.coelicolor/>) CDS are numbered using the following system eg SCB7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. Where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research. 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 301:157-66(1994) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 6A11 lies between and overlaps with cosmids 8F4 and 5D11 on the AseI-B genomic restriction fragment.
FEATURES
source location/Qualifiers
1. 28751
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 6A11"
2. 84
/note="Nominal overlap with Streptomyces coelicolor cosmid 5C6A11

/note="identified by Glimmer2, putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK25406.1"
 /db_xref="GI:13425162"
 /translation="MTLISAHKGCESGSPGFNAVLSSCGCYLAIFYAPCAQAPQRTPRPT
 WGRGRFHRRSSCTDPPRRSGRAAPPAQAPALPAPAPAAQAPADDAAPVPAALKEAATAPAA
 TVAPAPRGRTDPTFTVRLIDYLDVKYCKQVQAGDFQALSYGFEKKRKEQVPLAEI
 PRVATIDNGESNNVCTVITIDYQAPQEQALANGIDMATAWENSQPLALINDQTV
 SDPRRTYVWMDADMAAGRAGLIVRLMKRDKDETSGVKNFERAQLVSTSR"
 complement(1018..1278)
 /gene="CC3445"
 complement(1018..1278)
 /gene="CC3445"
 /note="identified by match to protein family HMX"
 /codon_start=1
 /transl_table=11
 /product="arv synthase FI, epsilon subunit"
 /protein_id="AAK25407.1"
 /db_xref="GI:13425163"
 /translation="NAKLHFSLVAPERELFSGEDVMQAPCAEGDFVLANHAPFMPTT
 LRGGVYVTKDGTIVTFDIOGGFADVGEGGLTILAEHVA"
 complement(1278..1529)
 /gene="CC3446"
 complement(1278..1529)
 /gene="CC3446"

gene
CDS

gene
CDS

CDS

gene

```

/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK25406.1"
/db_xref="GI:13425162"
/translation="MWIISAHKCGEGSPGVFNVLSSCGCYLAAAPCAARQNPTRPR
WGRSFRFRSSGNTNRSGSPAAPRQAPAPAPAPAAQAPADQAAAPCAAPAEAPAPAP
TYVAPVPGSPPTPTFTYRLDLIDKRCQVAGDGAOLVYSYGFKKRKEQVATLEIS
PENTYLDNPGSSNNKVCVCTIDYSQAEQAOELADHDAWENSPQRLILINDQTV
SDRRFTVSDDDMAGGRALVYMRKIKDETSGKNFERAQILVSTSR"
/complement(1018..1278)
/gene="CC3445"
/complement(1018..1278)
/gene="CC3445"
/note="identified by match to protein family HMW"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, epsilon subunit"
/protein_id="AAK25407.1"
/db_xref="GI:13425163"
/translation="NAKILFRSVAPRELFSESDVMQAPGAEQGVLANHAFMTT
LRGKATYVDGATTKTFVDIGQGFADVGVEGILLAEHAEVA"
/complement(1278..1529)
/gene="CC3446"
/complement(1278..1529)
/gene="CC3446"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK25408.1"
/db_xref="GI:13425164"
/translation="MTEDADALNCCDELAASVPDANAGIENDSLILINTGEREE
EGELTVLATIRFCFCEETQEDTDIEDALGAEVEETN"
/complement(1529..3318)
/gene="CC3447"
/complement(1529..3318)
/gene="CC3447"
/note="identified by match to protein family HMW"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, beta subunit"
/protein_id="AAK25409.1"
/db_xref="GI:13425165"
/translation="MATPEKPEPATAPKAPAPKAAAPKAAAPKAAAPKAPKAP
AAPPAAPAPAPAGGAGAGRLVQIIGAVDIEFGALPILINAVETYNITLGGVLE
FEVAHAGOSTVLAAMDAGEIVRQPDNGEPIIRVGEPTQIRINNVGEPIDE
OGPRLSISITIRHDAPEPEQNTNRELVLTGKILDLIDCKPTKGGKQLPGGAGGK
TVMQELINNIARAKYGYSTLACQVNGRTGEBDLHEHESNVNAPKANSGESGR
CAVYGMNEPPKARNAVALTGTSINAEYRDEBKGVLVFINIFRFTQAGAEVSLD
GRPSAQTPTLATNGNINQETISITNGSITSVOAIVPADDLDPAPATSFHLD
ATTVLSNDIAAKQIFPAVDPLSTSRIMDPLVGEHTVARVQGVLDQYKALDDII
AIGMDELSEDLVLAARAKRIORFLSQPFHVADEPTNPGAFVQDKDTIRSKYIVD
GEDHDPKPEAFVYGEIEHVAKAEKMAEA"
/complement(3196..4071)
/gene="CC3448"
/complement(3196..4071)
/gene="CC3448"
/note="identified by match to protein family HMW"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, gamma subunit"
/protein_id="AAK25410.1"
/db_xref="GI:13425166"
/translation="NASKIEKRNRRISVKAOTITKAMQWVAARAKLRSDQDAESAP
YAIRDLASVIANLAGVSGDGAPEKLLAGGRDRRLIYVAAAADRCLAGCFSSIVRAR
AIHIDLIOGDGVRVYCGKTKQIAKPYRGRITVEFDSSTKQTLVSAQPIADVI
TREVYAGTVDVLTLESFKSVQOQITGDLIPAVETGEAASGPAVAYEYEPSEA
ITTELITLIPNTVLTSLALDNMAQFVIAQMTADNATRNAGDMIKRTLEYNNSROA
ITKELEIETISAEANV"
/complement(4106..5638)
/gene="CC3449"

```

```

CDS      complement(4106..5638)
        /gene="CC3449"
        /note="identified by match to protein family HMM"
        /codon_start=1
        /transl_table=11
        /product="ATP synthase F1, alpha subunit"
        /protein_id="AAK25411.1"
        /db_xref="GI:13425167"
        /translation="MDIRAAEISALITKSOIANRGEAAVSDVGOVLSVSGDIANIYGL
        DNVOAGMELEFPKAGVGMALINERDVAVGATITRQDDIIEGDEVRNLGELVDVPGV
        GLGRVNPGEPIIDGKGPJOYTBERRADVAPGIIIRKSVHEPVQGLKSIDLIPY
        GRQREITIGDRQTKTAVADITLNRKANAGDESAKLYCVVAIGOKRSYAOIY
        KTLIEGALERTYVVAASAPAPLOYLAPGACGMEPEDNGLHGIITDDLSKQD
        VAYRQSLRLRPPGREGADYVYLHSLRLERAKINEDNGSGSLALPIITQDAD
        VSAIPLTNVISITDGIETDLEFYOGIRPAVAVNVSVSVGSAQIKAKQVAGPTK
        GEIAYREMAAFKFGSDLDASTOKMKARERLTELKOPYAPLSVEQCVIYAGT
        RGYLDGIPTSVSRFPEAFELARLHSHADLEIGIRTKALDKLENTLKSALDSFSSST
        FA"
        /complement(5691..6245)
        /gene="CC3450"
        /complement(5691..6245)
        /gene="CC3450"
        /note="identified by match to protein family HMM"
        /codon_start=1
        /transl_table=11
        /product="ATP synthase F1, delta subunit"
        /protein_id="AAK25412.1"
        /db_xref="GI:13425168"
        /translation="MADERKATPDAGCRYASQSLFELITENGSLQKVEADKSLKAMAD
        SADRLIASPFAESADKGLTAVAKKGFQPLTKGLLVANQNTGLDLAISAF
        VELSARHKGAVTREVSAALSPADKQVOTALQALGKPEVSTRVDPSSLGLKVR
        VSRLEFSDALRSLSLKFALPKRA"
        /complement(6338..6742)
        /gene="CC3451"
        /complement(6338..6742)
        /gene="CC3451"
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAK25413.1"
        /db_xref="GI:13425169"
        /translation="MDGRRCRAPSEANRLITRAFESRNQKATERSLPRSDGISPR
        RPPAPRRRGAVHAPARRSGSWVPVPPRCROGORSRPHRSRSARARRRPLRPCRE
        RLPVGRGSCSPVSTSPVAGKEKEAAGPAA"
        /complement(6831..7283)
        /gene="CC3452"
        /complement(6831..7283)
        /gene="CC3452"
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAK25414.1"
        /db_xref="GI:13425170"
        /translation="WLSDPKRRMGVRRRHPRGRASTRISPIIDLTATPKPEGFM
        KLRSLAAVALSVSMFAGSAPADGRIALADLPVAKRVVGVGAVFQCEGECVSAQ
        APSRLTAAACKALAKEVGRVAAFGCEKTSKDAEDLANQMSAKYAAK"
        /complement(7328..7864)
        /gene="CC3453"
        /complement(7328..7864)
        /gene="CC3453"
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAK25415.1"
        /db_xref="GI:13425171"
        /translation="MREAFSGIADLSPPAPRGMNRRLKPODLYDPALGIGVQATF
        DHALGAGVAAVLADPALDVGOGHSGRTGEGEGAHOGHGRADGCGGHRDRADPPF
        AEDLIQITDOKLAVRLILHHPDHDPEDGALDNRGQOHRHHAQNTKSDKQSEKCADRE
        GGHGDHAASVTPFMSRTG"
        /complement(7951..9258)
        /gene="CC3454"
        /complement(7951..9258)
        /gene="CC3454"
        /note="identified by match to PFAM protein family HMM"
        /codon_start=1
        /transl_table=11
        /product="acyl-CoA dehydrogenase family protein"
        /protein_id="AAK25416.1"
        /db_xref="GI:13425172"

```

```

CDS      /gene="CC3454"
        /gene="CC3454"
        /note="identified by match to PFAM protein family HMM"
        /codon_start=1
        /transl_table=11
        /product="acyl-CoA dehydrogenase family protein"
        /protein_id="AAK25416.1"
        /db_xref="GI:13425172"

Query Match
Best Local Similarity 5.6%; Score 99.2; DB 1; Length 14710;
Matches 213; Conservative 0; Mismatches 168; Indels 3; Gaps 1;

QY 56 CTTTGACCTGCTGCTCGCGGGGACCCCATCGACGAGGAGCAACACCCGGGGGGC 115
Db 14609 CCTAGCATCTGCTGCTCGCGGAGCGCGCTGCTGACGCGCGCGGCGGCTTCC 14550
QY 116 GCGCGGACCTGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 175
Db 14549 AGCGCGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14490
QY 176 CCGGCGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 235
Db 14489 GCGCGGAGGA--GATGACGCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14433
QY 236 ACACCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 295
Db 14432 ACACCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 14373
QY 296 GCGTCACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 355
Db 14372 GCGTCACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14313
QY 356 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 415
Db 14312 ATCAGCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14253
QY 416 CCGACTACTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 439
Db 14252 CCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14229

RESULT 15
AX211705
LOCUS AX211705 65140 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 1 from Patent WO0159126.
ACCESSION AX211705
VERSION AX211705.1 GI:15523937
KEYWORDS
SOURCE
ORGANISM Streptomyces noursei.
Streptomyces noursei.
Bacteria; Filicinutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 65140)
Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
Patent: WO 0159126-A 1 16-AUG-2001;
Norwegian Teknisk Naturvitenskapelige Universitet (NO); STRIETELSEN
IND OG TEKNISK FORSKNING VED NORGENS TEKNISKE HOGSKOLE (NO);
ALPHARMA AS (NO); Sluvenst AS (NO); Zotchev, Sergey Borisovich
(NO); Sekurova, Olga Nikolayivna (NO); Fjaervik, Espen (NO);
Brautaset, Trygve (NO); Stroem, Arne Reidar (NO); Valla, Svein
(NO)

FEATURES
source Location/Qualifiers
1..65140
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"

```

BASE COUNT 8270 a 25171 c 22273 g 9426 t
ORIGIN

Query Match 5.6%; Score 99.2; DB 6; Length 65140;
Best Local Similarity 43.5%; Pred. No. 0.0082;
Matches 724; Conservative 0; Mismatches 913; Indels 27; Gaps 5;

QY 62 ACCTGTGTGCGGGGCGGACCCATGAGACGAGCAACACCCCGGGGCGCGG 121
DB 19698 AGAGAGGCGCTGGCGGACGCTGCTTCAATGAGGCGCGCGCTTCCAGGCGCTGCGCGG 19757
QY 122 ACCTGGGCGTGGCGGCGGACCGGATCGCCGATCGGCGATGCTGTGAGAGCGCGCGCG 181
DB 19758 CTTGGGCGCGGACACGAGGCTTACGCGGAGTGGCCCTGCGCGGAGCGGAGACCG 19817
QY 182 AACCGCGGCTGACGCTGTGGGCGCTGTGTGCGCGCGCTTCACTGACGACCGC 241
DB 19818 ACCCGCGCGCTTGGAGTGCACCGCGCGCTGCTGAGCGCGGACACAGCGCGCGCT 19877
QY 242 ACAGGACACAACTACTGTCTACGCGGCGGATGACGCGGCAAGATCTGCGAGGCGTCA 301
DB 19878 AGCGGACCTGGCGGCGGATGACGCGGCGGCGGCTGCTGCTGCGGAGGCGTCTGCG 19937
QY 302 CCAGGCTGTACGCGGCGGATGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 361
DB 19938 TCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19997
QY 362 CCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 421
DB 19998 TCACGATGCGGCTTACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20057
QY 422 ACTGAGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 481
DB 20058 CCGCGGAGGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20117
QY 482 CAAAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 541
DB 20118 TCACGATGCGGCTTACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20177
QY 542 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 601
DB 20178 TCCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20237
QY 602 GCGGCTTGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 661
DB 20238 CCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20297
QY 662 CGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 721
DB 20298 TCACGATGCGGCTTACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20357
QY 722 TGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 781
DB 20358 CCGCGGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20417
QY 782 CGGACGACAAAGTCTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 841
DB 20418 TGGTCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20477
QY 842 TGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 901
DB 20478 GCGTGTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 961
QY 902 CCACGATGCTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 961
DB 20538 CCGAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20591
QY 962 AGCGCTTCCCGAAGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1021
DB 20592 ACAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20651
QY 1022 CCAAGTACGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1081

DB 20652 TCCGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20711
QY 1082 AACCCGACGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1141
DB 20712 GCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20768
QY 1142 CGGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1201
DB 20768 GCGACGCTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20828
QY 1202 ATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1261
DB 20829 CCGAATCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20888
QY 1262 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1321
DB 20889 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20939
QY 1322 TGGTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1381
DB 20940 TGGTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20999
QY 1382 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1441
DB 21000 GCGTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21059
QY 1442 TCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1501
DB 21060 GCGACGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21119
QY 1502 CTTTAC-----AATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1555
DB 21120 CCGGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21179
QY 1556 CAAGCGGTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1615
DB 21180 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21239
QY 1616 AGAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1672
DB 21240 TGACGACGAGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21299
QY 1673 AAGACTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1716
DB 21300 TGACCGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21343

Search completed: May 11, 2003, 03:03:11
Job time : 4976 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 11, 2003, 12:05:25 ; Search time 54.5 Seconds

(without alignments)
6201.993 Million cell updates/sec

Title: US-10-009-782-1
Sequence: 1 gaattccactgacgcgcga.....ccctgagctacagaagactt 1758

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODEL=frame+np.model -DEV=xip
-O=/cgsr2.1/USPTO.spool/US10009782/rumat.07052003-122517-23148/app-query.fasta.1.1927
-DB=PIR-73 -OPMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNIT5-BITS -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=US10009782 -ECGN 1.1.46 -gnat.07052003.122517-23148 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MAMP -LARGEORDER -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-NARN.TIMEOUT=30 -THEADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

PIR-73:
1: p1r1:
2: p1r2:
3: p1r3:
4: p1r4:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2511	76.1	484	2	JC4394	aminoacylase (EC 3
2	1000.5	30.3	488	2	JC4165	N-acyl-D-glutamate
3	942	28.6	526	2	B75202	d-aminoacylase (as
4	423.5	12.8	1106	2	J00405	hypothetical 119.5
5	407	12.3	581	2	B87678	hypothetical prote
6	338.5	10.3	529	2	T45134	hypothetical prote
7	337	10.2	924	2	S27923	gene lrf3 protein -
8	321.5	9.7	660	1	Q08E3	BHLH protein - hu
9	312	9.6	660	1	Q08E3	BHLH protein - hu
10	307	9.3	680	2	S21323	probable endogluca
11	296.5	9.0	1367	1	S48478	glucan 1,4-alpha-g
12	296	9.1	1791	2	T02345	hypothetical prote
13	284	8.9	611	2	D70928	hypothetical prote
14	289.5	8.8	3020	2	A43932	mucin 2 precursor,

C	15	287.5	8.9	1106	2	J00405	hypothetical 119.5
	16	284.5	8.6	1733	1	B45344	probable nuclear a
	17	282.5	8.6	437	2	C39135	hypothetical prote
	18	282.5	8.6	1460	1	EDBE1P	immediate-early pr
	19	277.5	8.4	494	2	G84348	hypothetical prote
C	20	274	8.4	1733	1	B45344	probable nuclear a
	21	273.5	8.3	1027	2	S28774	collagen alpha cha
	22	272.5	8.3	1414	1	S23809	collagen alpha 2(I
	23	270.5	8.2	507	2	T44768	antitrypsin glycop
	24	270	8.2	1464	2	GCHU15	collagen alpha 1(I
C	25	267.5	8.2	1958	1	B40505	collagen alpha 1(I
	26	267	8.1	825	2	J02345	DNA-binding protei
	27	265.5	8.0	1791	2	T02345	hypothetical prote
	28	265.5	8.0	3570	2	T45025	hypothetical prote
	29	264.5	8.0	1466	1	GCHU15	mucin MUC5B, trach
	30	264	8.0	1464	2	CGHD7L	collagen alpha 1(I
	31	263	8.0	1188	2	S59856	collagen alpha 1(I
	32	263	8.0	1453	2	S49915	collagen alpha 1(I
	33	262	7.9	1344	2	S21626	extensin-like prot
	34	262	7.9	1958	2	B40505	collagen alpha 1(I
C	35	261.5	8.1	640	2	T08179	mucin 1 precursor,
	36	261	7.9	779	1	CGBO1S	hypothetical prote
	37	261	7.9	839	2	F75518	collagen alpha 1(I
	38	260.5	7.9	580	2	T43451	hypothetical prote
	39	260.5	7.9	964	1	CGCH2S	probable mucin DKF
	40	260.5	7.9	1042	1	CGCH2S	collagen alpha 2(I
	41	258.5	7.8	2796	2	JC4743	collagen alpha 1(I
	42	255.5	7.7	1151	2	T18535	fatty-acid synthas
C	43	253	7.8	738	2	E87627	high molecular mas
	44	251	7.6	1419	2	A41182	hypothetical prote
	45	251	7.6	1487	2	B41182	collagen alpha 1(I

ALIGNMENTS

RESULT 1
JC4394
aminoacylase (EC 3.5.1.14) - Alcaligenes xylosoxydans subsp. xylosoxydans A-6
N: Alternate names: N-acyl-D-amino acid amidohydrolyase
C: Species: Alkaligenes xylosoxydans subsp. xylosoxydans A-6
C: Date: 20-Jan-1996 #sequence_revision 15-Apr-1996 #text_change 13-Sep-1998
C: Accession: JC4394
R: Nakayama, M.; Katsuno, Y.; Hayashi, S.; Miyamoto, Y.; Sakai, K.; Moriyuchi, M.
Biochem. Biotechnol. Biochem. 59, 2115-2119, 1995
A: Title: Cloning and sequencing of a gene encoding D-aminoacylase from Alkaligenes x
A: Reference number: JC4394; MIMD:96100942; PMID:851651
A: Accession: JC4394
A: Molecule type: DNA
A: Residues: 1-484 <NAK>
C: Comment: This enzyme, which catalyzes the hydrolysis of N-acyl derivatives of neur
residue of zinc ion or EDTA.
C: Genes:
A: Gene: dan
C: Superfamily: aminoacylase
C: Keywords: hydrolyase
F: 68-70/Region: zinc binding

Alignment Scores

Pred. No.: 1,64e-128
Score: 2511.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 76.11%
DB: 2
Gaps: 0

US-10-009-782-1 (1-1758) x JC4394 (1-484)

QY 34 ATGTCCCAATCCGATTCAGCCCTTCGACCTGCGGGCGGACCCGATCGAC 93
DB 1 MetSerGlnSerAspSerGlnProPheAspLeuLeuAlaGlyIleThrLeuIleAsp 20
QY 94 GGAGACACACCCCGGGGGGGCGGCGGACCTGCGGGCGGACCGACCGACCGCC 153

Db 21 GlySerAsnThrProGlyValArgValAlaAspLeuGlyValArgGlyAspArgIleAlaAla 40
 QY 154 ATGGCGATCTGTGAGAGCGCGCGGAGACACCCGGGTGACGTGTGGCGCTGGTGGTC 213
 Db 41 IleGlyAspLeuSerAspAlaAlaAlaIleThrArgValAlaValSerGlyLeuValVal 60
 QY 214 GCGCGCGGCTCATGAGCTGCACACCCAGCAGCAGCACTACCTCTCAGAGGTGGCGAC 273
 Db 61 AlaProGlyPheIleAspSerHisThrHisAspAspAspIleLeuAspArgArgAsp 80
 QY 274 ATGAGCGCCAGATCTGCAGCGGCTGCACACGGTGTGCACGGGCAATGGCGCATCAGC 333
 Db 81 MetThrProGlyIleSerGlnGlyValThrValValIleThrGlyAsnGlyIleSer 100
 QY 334 CTGGCGCGCTGGCGCAGCAGCCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
 Db 101 LeuAlaProLeuAlaIleAlaAlaAspProProAlaProLeuAspLeuAspGlyGly 120
 QY 394 TCTTACCGTTCGAGCGCTTGGCGACTACCTGAGCGCGTTCGCGGCGACCGCGCGCGC 453
 Db 121 SerTyrArgPheGlnArgPheAlaAspTyrLeuAspAlaLeuArgAlaThrProAlaAla 140
 QY 454 GTCAACGCGCGCTGTATGTGTGGCATTCACGCTGGCGCGCGCGCGCGCGCGCGCGCTG 513
 Db 141 ValAsnAlaAlaCysMetValGlyHisSerThrLeuArgAlaAlaValMetProAspLeu 160
 QY 514 CAGCGCGCGCGCGCAGCAGGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATG 573
 Db 161 GlnArgAlaAlaThrAspGlnGlyIleAlaAlaMetArgAspLeuAlaGlnGlyAlaMet 180
 QY 574 GCGCGCGCGCGCGCATTCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 633
 Db 181 AlaSerIleAlaIleGlyIleSerThrGlyAlaPheTyrProProAlaAlaArgAlaThr 200
 QY 634 ACCGAGAGATCATCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 693
 Db 201 ThrGlnGlnIleIleGlyValCysArgProLeuSerAlaHisGlyIleIleIleIleIle 220
 QY 694 CACATGGCGCGCAGAGCG 753
 Db 221 HisMetArgAspGlnGlyIleGlnHisIleValAlaAlaLeuGlnGlyIleIleIleIle 240
 QY 754 CCGGAGCTGAGCTGCGCGCGGTGATCTGCACACCAAGTATGAGCGCGCGCGCGCGCGC 813
 Db 241 ArgGlnLeuAspValProValAlaIleSerHisHisLysValMetGlyIleProAspPhe 260
 QY 814 GCGCGCTGCGCGGAGAGCGTGCCTGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
 Db 261 GlyArgSerArgGlnThrLeuProLeuIleGlnAlaAlaMetAlaArgIleAspValSer 280
 QY 874 CTGAGCGCTATCCCTACGTGCGCGCGCTGCACCATGCTCAACAGAGCGCGCGCGCGCGC 933
 Db 281 LeuAspAlaTyrProTyrValAlaGlySerThrMetLeuLysGlnAspArgAlaLeuLeu 300
 QY 934 GCGGAGCGACCATCATCATCTGTGTGCAAGCCCTCCCGGAACTAGCGCGCGCGCGCGC 993
 Db 301 AlaGlyArgThrIleIleThrTyrPysProPheProGlnLeuSerGlyArgAspLeu 320
 QY 994 GATGAGTGGCGCGGAGCGCGCAATCATCAAGTACGAGCGTGGTCCGCGCGCGCGCGCG 1053
 Db 321 AspGlnValAlaAlaGlnArgGlyIleSerIleTyrAspValValProGlnLeuLeuPro 340
 QY 1054 GCGCGCGCGCATCTTATGATGAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1113
 Db 341 AlaGlyAlaIleTyrPheMetMetAspGlnProAspValGlnArgIleLeuAlaPheGly 360
 QY 1114 CCGACCATGATCGGCTCGAGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1173
 Db 361 ProThrMetIleGlySerAspGlyLeuProHisAspGlnArgProHisProArgLeuTyr 380
 QY 1174 GGCACCTTCCCGCGGCTGCTGGGCGACTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1233
 Db 381 GlyThrPheProArgValLeuGlyHisTyrAlaArgAspLeuGlyLeuPheProLeuGln 400

QY 1234 ACGGCGATGAGATGACCGGCTGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1293
 Db 401 ThrAlaValThrPlysMetThrGlyLeuThrAlaAlaArgPheGlyLeuAlaGlyArgGly 420
 QY 1294 CAGTGCAGCGCGGCTACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1353
 Db 421 GlnLeuGlnAlaGlyTyrPheAlaAspLeuValAlaPheAspProAlaThrValAlaAsp 440
 QY 1354 ACG 1413
 Db 441 ThrAlaThrPheGlnHisProThrGlnArgAlaAlaGlyIleHisSerValTyrValAsn 460
 QY 1414 GCG 1473
 Db 461 GlyAlaProValThrGlnGlnGlnAlaPheThrGlyGlnHisAlaGlyArgValLeuAla 480
 QY 1474 CGCACG 1485
 Db 481 ArgThrAlaAla 484
 RESULT 2
 JCA165
 N-Acyl-D-glutamate amidohydrolase (EC 3.5.1.-) - Alcaligenes xylosoxydans subsp. x
 C:Species: Alcaligenes xylosoxydans subsp. xylosoxydans A-6
 C:Date: 12-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 07-May-1999
 C:Accession: JCA165
 R:Wakayama, M.; Ashtika, T.; Miyamoto, Y.; Yoshikawa, T.; Sonoda, Y.; Sakai, K.; Mo
 J. Biochem. 118, 204-209, 1995
 A:Title: Primary structure of N-acyl-D-glutamate amidohydrolase from Alcaligenes x
 A:Reference number: JCA165; MUID:96015170; PMID:8537313
 A:Accession: JCA165
 A:Molecule type: DNA
 A:Residues: 1-488 <MAX>
 A:Cross-references: DDBJ:D45918
 A:Note: The authors translated the codon CAG for residue 132 as Ala, GGC for resid
 C:Comment: This enzyme catalyzes the hydrolysis of N-acyl derivatives of various D
 C:Gene(s):
 A:Gene: dag
 C:Superfamily: aminocyclase
 C:Keywords: hydrolase
 Alignment Scores:
 Pred. No.: 8.09e-47 Length: 488
 Score: 1000.50 Matches: 223
 Percent Similarity: 59.38% Conservative: 62
 Best Local Similarity: 46.46% Mismatches: 190
 Query Match: 30.33% Indels: 5
 gaps: 4
 DB: 2
 US-10-009-782-1 (1-1758) x JCA165 (1-488)
 QY 52 CAGCCCTTCGACCTGCTGCTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 111
 Db 3 GlnLysLeuAspLeuValIleGlnGlyIleTyrPvalIleAspGlyLeuGlyIleProArg 22
 QY 112 CCG 171
 Db 23 ArgArgAlaAspValGlyIleArgGlyGlnArgIleAlaAlaIleGlyAspLeuSerAla 42
 QY 172 GCG 231
 Db 43 AlaProAlaAspArgAlaGlyLeuAlaGlyIleValAlaAlaProGlyPheIleAsp 62
 QY 232 TCGCACACCG 291
 Db 63 ThrHisGlyHisAspAspLeuMetPheValGlnLysProGlyLeuGlnTyrPheThrSer 82
 QY 292 CAGCGCGCTCACAGCGTGTCTACAGCGCAATTCGGCGATGAGCTGGCGCGCGCGCGCG 348
 Db 83 GlnGlyIleThrSerValValAlaGlyAsnGlyIleSerGlyAlaProAlaProLeu 102
 QY 349 CAGCGCAACCG 408


```

Db 196 TysSerLysThrTrpGluLeuIleGluLeuAlaMetValAlaAlaLysTyrGlyGlyLe 215
QY 685 TACGCCACCCACATGGCGCGGAGCGAGCCACATGCTGGCGCGCTGGAGAAACCTTC 744
Db 216 TyrSerThrHisMetArgSerPheGluGlyAsnArgValIleGluAlaLeuGluGluAlaIle 235
QY 745 CGCATGGCGCGCGAGCTGAGTGCAGTGCCTGGTATCTCCACACAAAGTCAATGGCCAG 804
Db 236 ArgIleGlyArgGluSerGlyAlaArgValIleGluValSerHisHisLysValSerGlyArg 255
QY 805 CCCAATTGGCGCGCTCGCGCGGAGAGAGCTGCGCTGATGAGAGCGCGCCATGGCGCGCAG 864
Db 256 LysAsnTrpGlyLysSerArgLysThrLeuAlaLeuIleGluLysAlaArgSngluGly 275
QY 865 ---GAGCTCTCGCTGAGACCGCTATCCCTACGTGGCGCGCTCCACATGCTCAAG----- 915
Db 276 IleGluIleThrLeuAspValTyrProTyrThrAlaGlySerThrTyrLeuAlaLeu 295
QY 915 ----- 915
Db 296 LeuProTyrValHisGluSerGlyLysIleLysGluArgCysArgAspGluThr 315
QY 916 -----CAGGACCGCGCTGCTGCTG 933
Db 316 ArgLysLysIleArgGluPheIleGluThrArgAspAspTrpGluAsnPheIleLysGlu 335
QY 934 GCCGGA-----CGCACCATCATCAGCTGTGCAAGCCCTTCCCGAAATGAGCGGGCGC 987
Db 336 AlaGlyTrpGluAsnIleIleIleThrHisSerGluAsnProGluPheValGlyLys 355
QY 988 GACTGATGAAAGTCCGGCGCGAGCGCGCAATCCAAAGTACGACGTGGTCCCGAGCTG 1047
Db 356 SerLeuLysGluIleSerAspLeuLeuHisArgSerProPheAspValLeuPheAspIle 375
QY 1048 -----CAGCGCGCGCGCGCATCTTCAATGATGAGCAACCGGAGCTG 1092
Db 376 LeuAlaLysAspGlyThrAsnAlaGlyMetIleValPheLeuMetSerGluGluAspVal 395
QY 1093 CAGCGCATCTGCTGGCTCGCGCGGACCATGATGGCTCCGACCGCTCGCGCGAG 1152
Db 396 GluArgIleLeuSerHisProTyrSerMetIleGlyThrAspGlyLeuAspSerGlyLys 415
QY 1153 ---CGCGCGCATCGCGCTGTGGGCGACCTTCCCGCGGCTGTGGGCGCATTTAGGCGC 1209
Db 416 GlyLeuProHisProAlaArgIleGlyThrPheProAlaValLeuGlyArgTyrValArg 435
QY 1210 GACTGGCGCTGTCCCGCTGGAGACGGCGGTATGGAAGATGACCGCGCTGACCGCGCG 1269
Db 436 GluLysLysLeuLeuArgLeuGluAspAlaIleArgLysMetThrSerLeuProAlaLeu 455
QY 1270 CGCTTGGCTGCGCGCGCGCGCGGAGCTGCAAGCGCGGCTACTTGGCGAGCTGTGTG 1329
Db 456 LysLeuGlyLeuLysAspArgIleValLysGluGlyMetThrAlaAspLeuValIle 475
QY 1330 TTCGACCGCGCGCGAGTGCAGTACCGCGCATCTTGAACACCTTACGAGCGCGCGCGC 1389
Db 476 PheAspProHisArgValLysAspArgAlaThrTyrThrAsnProArgLeuProAsp 495
QY 1390 GGCATTCATTCGTAGTGTCAACGGCGCGCGCTGTGCAAGACGAGCGCTTACCGCGC 1449
Db 496 GlyIleLysTyrValIleValAsnGlyValLeuSerValGluAsnGlyLysLeuThrGly 515
QY 1450 CAGCATGGCGCGCGCTGCTGCGACGCGAGCGC 1482
Db 516 AspAlaGlyGlyValIleArgArgThrSer 526

```

RESULT 4

J00405
 hypothetical 119.5K protein (uvrA region) - Micrococcus luteus
 NAlternate names: ORF 1 protein
 C:Species: Micrococcus luteus, Micrococcus lysodeikticus
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000

```

C/Accession: J00405
R:Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A:Title: Micrococcus luteus homolog of the Escherichia coli uvrA gene: Identifica
A:Reference number: S04781; MUID:89364717; PMID:2549377
A:Accession: J00405
A:Molecule type: DNA
A:Residues: 1-1106 <SH1>
A:Cross-references: EMBL:X15867
A>Note: this reading frame extends between two stop codons and does not begin wit
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal hc

```

Alignment Scores:

Pred. No.:	1-4e-15	Length:	1106
Score:	423.50	Matches:	239
Percent Similarity:	32.84%	Conservative:	28
Best Local Similarity:	29.40%	Mismatches:	252
Query Match:	12.84%	Indels:	294
DB:	2	Gaps:	47

US-10-009-782-1 (1-1758) x J00405 (1-1106)

```

QY 38 CCCAATCGATTCACAGCCCTTTCGACCTGTGCTGCGCGCGGCGACCTCATCGAGCGCA 97
Db 42 ProValSerCysProCysProCysProLeuCysTrpProAlaGla-----GlnGluLeu 58
QY 98 GCAACACCCCGCGCGCGCGCGCGCGAGCTGGCGCTGCGCGCGAGCCGAT---CGCGCGCA 154
Db 59 Leu-HisHisArgLeuLeuArgArgArgGlyProArgArgArgProArgLeuArgProAr 78
QY 155 TCGGCGATGTGTGAGAGCGCGCGCGCGCGACAC-----CGGGGTGA---CGTGTGCGGC 205
Db 78 GArgArgAlaLeuLeuLysArgAlaGlyProHisArgArgProGlyArgProArgAlaGlnPr 98
QY 206 TGGTGTGCGCGCGCGCGCTTATGACATGCGACACCCAGCAACAACTACTGCTGAGC 265
Db 98 OGluGlyArgGly-----CysGlnLeuPro-----Al 107
QY 266 GTGCGCATGACGACCCCMAGATCTCCAGGCGCTCACACAGTGTGT-----CACGGCGCA 319
Db 107 agLArgGHis-----GlyArgValHisGlySerValArgLeuGly 121
QY 320 A-----TTCGGGCT-----CAGCTGGCGCGCGCTGGCGCGACCGCGCGC 361
Db 121 nValLeuProGlyLeuAlaThrHisAspLeuArgArgGlyProAlaAlaLeuArgValAl 141
QY 362 C-----CGCGCC-----CCTGACCTGTGAGAGCGAGCGCGCTTACCGTTTCG 406
Db 141 aleuLeuLeuArgProHisValArgGlyProGlyLysGlnAlaGly-----Ar 157
QY 407 AGCGCTGCGCGA-----CTACCTGAGCGCTT----- 434
Db 157 gelyLeuHisArgGlyProValArgGlyArgValHisArgProGluValHisGlnProG 177
QY 435 ---CGGGCGACGCGCGCGCGCGCGCTCAAGC----- 461
Db 177 nProAlaLeuHisSerGlyAspHisHisArgSerLeuArgLeuHisAlaProAlaLeuG 197
QY 462 -----CGCTGTATGAGTGGCGCATTCAC---GC 487
Db 197 YThrCysArgGlyAlaAlaLeuProAlaValAlaArgArgAlaGlyGluProAlaAspProAl 217
QY 488 TGGCGCGCGCGCTCATGCGCGAGCTTGCAGCGCGCGCGCAC-----CG 529
Db 217 aAlaAspArgGlyProAlaArgGlyAlaAlaArgAlaHisProLeuProGlyAlaArgAl 237
QY 530 ACAGGAATCGCGCGCATGCG-----GAGACCTGCGCGAGAG-----ACCATGG 574
Db 237 aArgGlyProArgProGlnGlyArgValArgGlyProValGlnGlyProValHisAlaG 257
QY 575 CGAGCGCGCGCATTCGACCGCGCGCTTCTA---CGCGCGCGCGCGCGCGC--- 629

```


Db 6 TyrAspLeuValLeuArgSerGlyValValAlaPheAspGlyAlaGlyProGlyPheGln 25
 QY 118 GCCACCTGGGCGTGGCGGAGCGACGATCGCGGCATCGCGCATCTGTGCGAGCGGCC 177
 |||||
 Db 26 AlaAspValGlyValAlaArgAspGlyValAlaValAlaValGlyLys---GlyLeuAlaAla 44
 QY 178 GCGCACACCCGGTGCAGCTGTGCGCGTGTGCGCGCGCGCTTCATGACTGCGAC 237
 |||||
 Db 45 GlyAlaGlnGluLeuAspAlaArgGlyAlaArgLeuAlaThrProGlyPheValAspIleHis 64
 QY 238 ACCACACGACGACACTACCTGCTAGCGGTGCGCATGACGCCCAAGATCTGCGAGGC 297
 |||||
 Db 65 ThrHisTyrAspGlyGlnAlaThrTrpSerGlyGlnLeuGlyProSerSerGlyHisGly 84
 QY 298 GTCACACGAGTGTGCGCGCATTTGGGGCATCGCGCTGGCGCGG----- 342
 |||||
 Db 85 ValIleThrValValIleMetGlyAsnGlyValGlyIlePheAlaProCysAlaProAsp 104
 QY 343 -----CTGGCGCAGCCACCCCGCGCGCGCGCGCGCGCTGCGACTGTGCGAC 384
 |||||
 Db 105 HisAspArgLeuIleArgLeuMetGluGlyValGluAspIleProPheProValLeuThr 124
 QY 385 GAAGGGCGGCTCTACCTTCGAGCGCTTGGCGCATCTGAGCGGCTTGGCGCGCAGC 444
 |||||
 Db 125 GluIleGlyLeuProTrpAlaTrpGluSerPheProAspTyrLeuAspPheLeuAlaGlyArg 144
 QY 445 CCGCGCGCGCTCAACCGCGCTGATGTGGCGCATCAACCGCGCGCGCGCGCGCTGATG 504
 |||||
 Db 145 AlaPheAspValAlaAspValGlyAlaGlnLeuProHisAlaAlaLeuAlaValTyrValMet 164
 QY 505 CCGGACTTG-----CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAC 555
 |||||
 Db 165 GlyAspArgGlyAlaAspArgGluProAlaThrAspAlaPheAlaIleAlaMetAlaAla 184
 QY 556 CTGGCGCGGAGAACGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTAC--- 612
 |||||
 Db 185 IleAlaArgArgAlaValGluAlaGlyAlaLeuGlyPheSerThrSerArgThrLeuAsn 204
 QY 613 -----CCG 648
 |||||
 Db 205 HisArgThrSerAspGlyGlnProThrProThrLeuThrAlaGlyGluAspIleLeuThr 224
 QY 649 GAGGTGCGAC--- 705
 |||||
 Db 225 GlyIleAlaLeuGlyLeuAlaAlaAlaGlyGlyValLeuGlnValValSerAspPhe 244
 QY 706 ---GAAGGGGAGACATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 759
 |||||
 Db 245 ValAspGlyAlaAlaGluLeuAlaMetLeuArgArgIleValGluArgSerGlyArgPro 264
 QY 760 CTGGAGCTGCGCGGTGTGATCTCG----- 783
 |||||
 Db 265 LeuSerPheSerLeuValGlnSerProIlyAlaProGluGlyTyrPArgIleLeuAsp 284
 QY 784 -----CACCACCAAGGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
 |||||
 Db 285 GlyValAlaAlaAlaValAlaAspAlaGlyLeuProMetGlyAlaGlnValCysGlyArgPro 304
 QY 808 -----AATTGCGC----- 816
 |||||
 Db 305 ValGlyValLeuPheGlyLeuGluLeuThrLeuAsnProPheSerGlnAsnProValPhe 324
 QY 817 CGCTCGCGCGGAGCGCTGCGGTGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 876
 |||||
 Db 325 AlaGluLeuLysAspArgProLeuAlaAspLysValAlaAlaLeuSerAspProAlaPhe 344
 QY 877 GACCGCTATCCCTAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 927
 |||||
 Db 345 ArgAla-----ArgLeuLeuAlaHisAspIlyAspAlaArgGly 357
 QY 928 CTGCTGCGCGGAGCGACATCATCACTGTGCGAGCGCGCTTCCCGCACTGAGCGCGCG 987
 |||||
 Db 358 ProPheAlaGlySerAlaLeuArgAlaTrpAspAsnLeuTyrPro---MetGlyGlyVal 376

QY 988 GACCTGAT-----GAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1017
 |||||
 Db 377 AspProAspTyrGluProThrAlaAspArgThrValAlaAlaIleAlaAlaArgGluGly 396
 QY 1018 AATATCAAGTACGACGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1065
 |||||
 Db 397 ArgAspProAlaAlaValValLeuAspAlaMetLeuAlaArgAspGlyArgGlyMetLeu 416
 QY 1066 TACTTC---ATGATGAGCAAGCCGAGCTGCGAG-----CGCATCTGCGCG 1107
 |||||
 Db 417 TyrHisProPheLeuAsnTyrAlaAspGlySerLeuAspProSerPheAlaMetLeuSer 436
 QY 1108 TTGCGCGCGGACCATGATCGCG---TCCGAGCGCGCTGCGCGAC-----GAC 1149
 |||||
 Db 437 HisArgAspThrValProGlyLeuSerAspGlyIlyAlaHisValGlyMetIleCysAsp 456
 QY 1150 GAGCGCGCGGATCG 1206
 |||||
 Db 457 -----GlySerPheProThrSerAsnLeuIleHisTrpThr 468
 QY 1207 CCGGACCTG-----GGCTGTTCGCGTGGAGACGCGCGGTATGGAAGATGACCGGC 1257
 |||||
 Db 469 ArgAspArgThrArgGlyProArgIleProLeuGlyThrMetIleAlaArgGlnSerArg 488
 QY 1258 CTGACCGCGCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1317
 |||||
 Db 489 AspThrAlaGlnAlaValGlyLeuLeuAspArgGlyLeuIleAlaProGlyTyrArgAla 508
 QY 1318 GACCTGTGTGTTTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1368
 |||||
 Db 509 AspLeuAsnValIleAspTyrAlaGlyLeuArgGluAlaProGlnValAlaTyrAsp 528
 QY 1369 CACCT-----ACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1410
 |||||
 Db 529 LeuProAlaGlyGlyArgArgLeuThrGlnArgAlaHisGlyTyrValAlaThrIleVal 548
 QY 1411 AACGGCGCGCGCGCTTGGCAAGACGAGCGCTTACCGCGCGCGCGCGCGCGCGCGCG 1470
 |||||
 Db 549 AlaGlyValValThrGlnArgAspGlyGluProThrGlyAlaLeuProGlyArgLeu--- 567
 QY 1471 GCACGACGCGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1530
 |||||
 Db 568 -----ValArgGlyAlaGln 572
 QY 1531 TGGCGCGCGCTC 1542
 |||||
 Db 573 SerAlaProLeu 576
 |||||

RESULT 6
 T45134
 Hypothetical protein [imported] - Microbacterium ammoniaphilum (fragment)
 C:Species: Microbacterium ammoniaphilum
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T45134
 R:Striebel, H.M.; Seebert, S.; Jarsch, M.; Kessler, C.
 Gene 172, 41-46, 1996
 A:Title: Cloning and characterization of the Mami restriction-modification system
 A:Reference number: 222923; MID:96257250; PMID:8654988
 A:Accession: T45134
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-529 <STR>
 A:Cross-references: EMBL:X79027; NID:9984667; PIDN:CA55649.1; PID:g1679831
 A:Experimental source: ATCC 15354

Alignment Scores:
 Pred. No.: 5,19e-11 Length: 529
 Score: 338.50 Matches: 181
 Percent Similarity: 35.59% Conservative: 45
 Best Local Similarity: 28.50% Mismatches: 198
 Query Match: 10.26% Indels: 211
 Db: 2 Gaps: 30

US-10-009-782-1 (1-1758) x T45134 (1-529)

```

OY 32 CCAATGCCAATCCGATCCCGACCTTCG---ACCTGCTGCTGCGG-----76
Db 1 ProAlaProSerAlaThrArgSerProSerTrpSerIleCysSerArgIleProThr 20
OY 77 -----GGGCGACCTCATCG 91
Db 21 SerProAlaThrThrSerSerThrCysArgThrAlaProAlaAlaSerAlaProAlaSer 40
OY 92 ACCGCAACAAACACCCCGGCGCGCGCGACCTGCGCTGCG-----134
Db 41 ThrAlaGluProHisAlaGluProAla-LysProGluArgAlaGluProArgGluThrG1 60
OY 135 -----CGGCGACCGCATCGCGCGCATCGCGCATCTGTGTGCGACGCGCGCGC 181
Db 60 ymTrIleGluProArgGlySerHisTrpArgHisArgArgProAlaGluArgGluGlnVa 80
OY 182 ACACCGGCGTGCACGTGCTGCGCGCTGCTGCGCGCGCTCATCGACCTCGACACCC 241
Db 80 HisProLeuGlnArgPro-----HisGlnGluProGluAl 92
OY 242 ACAGACACAACTACTCTGCTCAGCGCTGCGCATGACGCCCAAGATCTCGACGGCGCTCA 301
Db 92 aaArgGlyGluLeuProValArgAspAspArgAlaGlu-----104
OY 302 CCACGGTGTGCACGGGCATTTGGCGCATGACCTGGCGCGCTGCGCGCACGCCACCCGC 361
Db 105 -----ArgGlyArgArgGluProAl 111
OY 362 CCGCCCCCTGGAACCTGCTGAGACGAGCGGCTTTACCGTTTCGAGCGCTGCGCGACG 421
Db 111 aaArgProAlaAlaArgAspAlaArgArgHisLeu-----122
OY 422 ACCTGAGCGCTGCGCGCGCGCGCGCGCGCTCAACGCCCTGTATGCTGCGCAT 481
Db 123 -----ArgLeuGlyAlaHis-----127
OY 482 CAACGCTGCGCGCGCGCGCTCATGCGGATGCGCGCGCGCGCGCGCGAGAAATCG 541
Db 128 ----ProAlaArgArgGly-----ValLeuArgArgHisArgArg---HisArg 141
OY 542 CCGGCATGCGGGAACCTGCGCGGAGAAAGCCATGCGCGCGCGCATTCGATTCGACCG 601
Db 141 gAlaArgGlyGluArgGlyArgGly-----ProArg 151
OY 602 GCGGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661
Db 151 gGlnGlnValProArgGlnHisProArg---GlyArgArgAspArgAlaGlyArgProG1 170
OY 662 CGCTGAGCGCGCATGCGCGCATCTACGCGACCGACCGCGCGCGCGCGCGCGCATCG 721
Db 170 yLeu-----HisArgArgArgArgArgAlaArgArg 180
OY 722 TGCGCGCGCTGAGAAACTT-----CGGCATCG 751
Db 180 gArgGlnGluArgProGlnValArgHisGlyAspArgGlnHisArgAlaAspArgArg 200
OY 752 GCGCGGAGCTGAGCTGCGCGCTGCTATCTGCACACAGAGTCATGCGCGCGCGCGCG 811
Db 200 gProArgAspProArgAlaGluHisProLeuArgGluGluGlnGlnGlnGlnAla 220
OY 812 TGCGCGCGTGCAGGAGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 871
Db 220 aaArgProLeuArgAla-----ArgGlyGlyGlnGlnGlyProGluGlyArg 235
OY 872 CGCTGAGCGCGTATCCTACCTGCG-----CGGCTCCACCATGCTCAAGAGCGCGCG 925
Db 235 caLaGlyArgHisProAlaGluGlyAspArgProArgAlaArgGlnGlnGlnAla 255
OY 926 TGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 985

```

```

Db 255 gAlaAlaHisArgIleAlaLeuHisLeuArgLeuGlnArg-----268
OY 986 GCGACCTGAGTGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1045
Db 269 -----ArgArgGlyArgAla-----AspIleCysArgAl 278
OY 1046 TGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1105
Db 278 aGlnGlyArgAlaArg-----ArgAlaArgArgAlaGlyGlyGly 292
OY 1106 CGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1141
Db 292 yLeuProArgArgGluArgValArgAlaGlnHisArgProArgProArgGlyG1 312
OY 1142 CGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1201
Db 312 yAlaAlaArgLeuAspArgAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyG1 332
OY 1202 A---TGCGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1232
Db 332 uArgHisProArgProAlaArgLeuProHisGlyGlyProGlnGlyValAlaArgLeuAs 352
OY 1233 ---GACGCGCGTATGGAAGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1288
Db 352 rHisProGlnGlyLeuGlnGlyGlyAlaAlaGlySerArgArgHisProHisArgLeuArgAl 372
OY 1289 GCGGCGAGCTGACGCGCGG-----GTACTTGCAGCACT-----1322
Db 372 aaArgLeuHisGlnGlyArgGlyAspLeuAlaArgProArgArgArgArgLeuGlyArg 392
OY 1323 -GGTGTGTTGCACCGCGC-----CACGCTGCGCGCATACCGCGCGCGCGCGCG 1372
Db 392 gArgGlyProArgGlnGlyGlnGlyAlaHisGlyGly-----GlnGlyLeuArgHisAl 410
OY 1373 CTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1432
Db 410 aGlyArgArgArgArgGlyGlyAlaProSerGlnLeuAlaValLeuArgGlnGlyGlyAla 430
OY 1433 AGCAGCGCTTCAC-----GGCGACCATGCGCGCGCGCGCGCGCGCGCGCGCG 1483
Db 430 nAlaSerAla-ThrLeuAlaLeuTrpMetThrSerGlyArgLeuLeuArgGlyLeuAla 450
OY 1484 CCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1543
Db 450 yLeuLeuCysArgAspSerValGluSer-----459
OY 1544 CAACCTGAGCAACCAACCGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1600
Db 460 --HisLeuAspValThrThrArgAlaAspAlaSerIleSerArgArgThr 476
OY 1601 CCGATATCGTGGCGCAAGAGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1656
Db 477 -----TrpThrValMetAsnThrArgLeuArgGlnAlaGlnLysSerLeuA 492
OY 1657 -----CGGAAATGACCTGCAAGACCTGTGCGACGCG 1689
Db 492 IaAlaArgArgProThrThrSerSerSerIleSerArgSer 505

```

RESULT 7
 527923
 gene Lf3 protein - human herpesvirus 4
 C/Species: human herpesvirus 4, Epstein-Barr virus
 C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Jan-2000
 C/Accession: S27923
 R/Parker, B.D.; Bankier, A.; Satchwell, S.; Barrett, B.; Farrell, P.J.
 submitted to the EMBL Data Library, August 1990
 A/Description: Sequence and transcription of Raj1 Epstein-Barr virus DNA spanning the
 A/Accession: S27923
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-924 <PAR>
 A/Cross-references: EMBL:M35547; NID:g330420; PIDN:AAA45896.1; PID:g330421

C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Alignment Scores:

Score: 6.58e-11 Length: 924
Percent Similarity: 33.700 Matches: 210
Best Local Similarity: 34.588 Conservative: 30
Query Match: 10.226 Mismatches: 287
Indels: 168
Gaps: 31

US-10-009-782-1 (1-1758) x S27923 (1-924)

```

QY 45 CGATGCCAGCCCTTCGACCT---GGTCTCCGGGGGGCCGACCTTCATCGACGGACGAA 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 ArgGlyProCyLeuAlaProGlySerGlyLeuGlyAlaHisProHisProArgSer 25
QY 102 -----CACCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26 GlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProAlaProGly--- 44
QY 132 GCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 45 ---ProGluProArgThrArgLeuInProAlaThrProArgSerGlyAlaAlaAsp 63
QY 180 -----GCACACCCGGGT-----CGACGTGGGGGGGGGGGGGGGGGGGGGGGGGG 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 ProAlaAspProValGlyHisProAlaAlaProAlaProGlyProGluProArgThr 83
QY 219 CGGCTTCATCGACCTCGACACCCACAGACAGACACTACTGCTGCGGGCGGCGGCGGCGG 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 ArgLeu---GlnProAlaThrProArgSerGlyAlaAlaAspProAlaAspProVal 102
QY 279 GCCCAAGATCTCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103 GlyHisProAlaAlaProGlyAlaProGlyProGluProArgThrArgLeuInProAla 122
QY 339 GCC-----GCTGGCGGACGGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 ThrProArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaPro 142
QY 369 -----CCTGACCTGTGGACGAGCGGCTC-----TTACGTTTCGAGCGCT 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 ArgAlaProGlyProGluProArgThrArgLeuInProAlaThrProArgSerGly 162
QY 414 CGCGCACTACCTGAGCGCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 AlaAlaAspProAlaAspProValGlyHis---ProAlaAlaProAlaProGlyProG 182
QY 474 GGGCGCATTCAGCGCTCGCGCGCGGCTCATGCGGACTTGCAGCGCGCGCGCGCGCGAGCA 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 LurProArgThrArgLeuInProAlaThrProArgSerGlyAlaAlaAspPro--- 200
QY 534 GGAATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 ---AlaAspProValGly-----HisProAlaAlaProAlaProGlyAla 213
QY 594 TTCGACGGCGGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 213 ToglProGluProArgThrArgLeuInProAlaThrPro---ArgArgSerGlyAlaAla 232
QY 642 -----GATCATCGAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 AspProAlaAspProValGlyHisProAlaAlaProAlaProGlyProGlu 250
QY 693 CCACATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 ProArgThrArgLeuInProAlaThrProArgSerGlyAlaAlaAspProAlaAsp 270
QY 753 CGCGGAGTGAAGTGGCGGCTGATCTGCGACACAGGATCATGGGCGGCGGCGGCGGCGGCGG 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 Pro-----ValGlyHisProAlaAlaProAlaProGlyProGluProArg 286
QY 804 GCCCAATTTCGGCGGCGGCGG-----CGAGCGCTGGCGGCGGCGGCGGCGGCGGCGGCGG 848

```

```

Db 287 ThrArgLeuInProAlaThrProArgSerGlyAlaAlaAspProAlaAspProVal 306
QY 849 CGCCATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 GlyHisProAlaAlaProGlyAlaProGlyProGluProArgThrArgLeuIn---Pro 326
QY 900 CTCACCATGCTCAAGCAGACCGCGG-----TGCTGCTGGCGGCGGCGGCGGCGGCGG 944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 326 LurProArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaPro 346
QY 945 CATCATCACTGGTGCAGAACCTTCGCCGAATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1004
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 RohGlyAlaProGlyProGluProArgThrArgLeuInProAlaThr-----ProArg 364
QY 1005 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 rgsSerGlyAlaAlaAspProAlaAspPro---ValGlyHisProAlaAlaProAlaPro 383
QY 1059 -----CGCATCTACTTCATGATGAGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGG 1106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 GlyProGluProArgThrArgLeuInProAlaThrProArgSer-----Gly 400
QY 1107 GTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 AlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyProGlu 420
QY 1167 CTTGTGGGCGACCTT-----GCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ProArgThrArgLeuInProAlaThrProArgSerGlyAlaAlaAspProAlaAsp 440
QY 1182 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 ProValGlyHisProAlaAlaProArgAlaProGlyProGluPro--- 455
QY 1242 ATGGAAGATGACGGCGCTGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 -----ArgThrArgLeuInProAlaThrProArgSerGly 468
QY 1302 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 AlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyProGlu 488
QY 1361 CCTTGGAACACCTA----- 1375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 ProArgThrArgLeuInProAlaThrProArgSerGlyAlaAlaAspProAlaAsp 508
QY 1376 -----CGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 509 ProValGlyHisProAlaAlaProArgAlaProGlyProGluProArgThrArgLeuIn 528
QY 1424 TCTGGCAAGACAGCGGCTTCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 ProAlaThrProArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAla 548
QY 1466 TGCTCGCGAGACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 AlaProArgAlaProGlyProGluProArgThrArgLeuInProAlaThrPro---Arg 568
QY 1520 ACGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 568 gSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaPro 587
QY 1565 ACATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 588 ---GlyProGluProArgThrArgLeuInProAlaThrProArgSerGlyAlaAla 606
QY 1625 T-----GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 AspProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyProGluProArg 626
QY 1665 -----GACCTGCAAGACCTGCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

Db 546 -----ProProglYalaglyGlnarProSerGlyProThr-----GlyGlyArProAla 563
QY 1536 CCCCCTCCCAACCTGGAGCAACCGCTCAATGAGCC-----CTCCCTCCG 1583
Db 563 labProglYalArProglYThrProAlaAlaProglYProglYgLyAlaAlaValProS 583
QY 1584 TCCCAATACGGCCCAACCGCATTCGTGGCAAGAGAGATGGGGCCGCTGGCCG 1643
Db 583 etGlyAlaThrProHisPro-GlnArGlySerGlyProAlaAsProProAlaAla 602
QY 1644 CGAGGCCAAGCCCGAATAATGACCTGCAAGACCT-----GTCCAGGCCAGCGG 1694
Db 603 ArgLeuProProglYArGlnGlnGlnProArgLeuProGlnAspLeuAlaAlaGlnAr 622
QY 1695 ---CATCCGGCTCTCGACCGCTGTCAGAGCCGACGTGGCCGACGATCCGCTGAGCTAGA 1751
Db 623 CysProAlaGlyProProProThrArGSerGlyAlaAlaAlaGlnArGThrHisArG 642

```

RESULT 9

BHFL1 protein - human herpesvirus 4 (strain B95-8)
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C:Date: 25-Feb-1985 #sequence, revision 25-Feb-1985 #text, change 23-Aug-1997
 C:Accession: A03742
 R:Bankier, A.T.; Delinger, P.L.; Farrell, P.J.; Barrell, B.G.
 Mol. Biol. Med. 1, 21-45, 1983
 A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
 A:Reference number: A93065; MUID:85055713; PMID:6092825
 A:Accession: A03742
 A:Molecule type: DNA
 A:Residues: 1-660 <BAN>
 R:Bankier, A.T.; Biggin, M.D.; Delinger, P.L.; Farrell, P.J.; Gibson, T.J.; H
 Nature 310, 207-211, 1984
 A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
 A:Contents: annotation; protein coding region
 C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
 C:Superfamily: human herpesvirus 4 BHFL1 protein

Alignment Scores:

```

Pred. No.: 1,44e-09 Length: 660
Score: 312.00 Matches: 181
Percent Similarity: 31.59% Conservative: 30
Best Local Similarity: 27.10% Mismatches: 232
Query Match: 9.62% Indels: 225
DB: 1 Gaps: 31

```

US-10-009-782-1 (1-1758) x Q0BE3 (1-660)

```

QY 1753 TCTCGAGCTGAGCGCATCTGCCACGCT-----CGGGCT 1718
Db 87 SerAlaGlnSerAlaArgArGThrGlyProAlaGlnAlaAspHisAlaHisSerAsnPro 106
QY 1717 TGGACAGGATCGAGACCGCATGCCGCGCTGGCGCTGACACAGCTTCGAGGGTCAATTTC 1658
Db 107 ThrGlyGlyCysSerAspProGlnArGSerProArGThrArGlnAlaGlyAlaValAla 126
QY 1657 GGGCTTCGCTCGCGCGGAGCGCGCGCCATCACTTCCTGCGCCACGATACGGGTG 1598
Db 127 GlyGlnGlySerAlaGlyLeuGlySerArGlyProArGProHisProAlaPheGlnVal 146
QY 1597 -----GGCGCGTATTGC----- 1586
Db 147 GlnTrpSerAlaArGAsnProGlyCysProArGThrTrpArGArGArGSerGlyAlaGln 166
QY 1585 -----GAGCGAGGAGGAGGCGCCATGATACGGT----- 1559
Db 167 ArgGlyHisProProProGlyAlaGlyGlnArProSerGlyProThrGlyGlyArGPro 186
QY 1558 -----TTGCGTCCAGGCGTGGAGGGGCG----- 1535
Db 187 AlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaVal 206

```

```

QY 1534 -----GGCAGCCGCCCGCTTCACCCGAGATGTAAAGGCTGGCGCGGCTCAGCGG 1481
Db 207 ProSerGlyAlaThrProHisProGlnArG-----GlySerGlyProAlaAsp 222
QY 1480 CCGTGGCTCGAGCAAGC----- 1457
Db 223 ProProAlaAlaAlaArGLeuProProGlnArGlnGlnProArGLeuProGlnAspLeu 242
QY 1456 CATGCTGCGCGGTGAACCGCTCTGTCGACAGCGGG----- 1418
Db 243 AlaAlaAlaGlnArGArGProAlaAlaGlyLeuProProProThrArGSerGlyAlaAlaGln 262
QY 1417 CGCGCTTACGTACACGGAATGATGCCGCGCGCGCGCTGATAGGCT-----GTGGA 1364
Db 263 ArGThrHisArGArGProProGlyCysProArGSerAlaArGAsnProGlyCysArGAr 282
QY 1363 AGTGGCGGTATGCG----- 1328
Db 283 ThrTrpArGArGArGSerGlyAlaGlnArGlyHisProProProGlyAlaGlyGlnArG 302
QY 1327 CCACCAAGCTGCGCGCAATACCCGCGCTGACAGT-----GCCCGCGCGG 1283
Db 303 ProSerGlyProThrGlyGlyArGProAlaAlaProGlyAlaProGlyThrProAlaAla 322
QY 1282 CCAGCGCGAAGCGCGCGCGGTCAATCTTCATCCAGCGCGCTCCAGCGGGA 1223
Db 323 ProGlyProGlyGlyGlyAlaAlaAlaProSerGlyAlaThrProHisProGlnArGly 342
QY 1222 ACAGGCCAAGCTGCGCGCATACAGCCCA----- 1193
Db 343 SerGlyProAlaAspProProAlaAlaAlaArGLeuProProGlnArGlnGlnProArG 362
QY 1192 -----GCACCCGCGGAGAGTGGCCCGACAGCGCGGATGCGGCGCTCGT 1148
Db 363 LeuProGlnAspLeuAlaAlaGlnArGArGProProProThrArGSer 382
QY 1147 CGTGGCGAGCGCGCGCGGAGCGCGATCATGTGTGGCCGACCGACAGATGCGCTGACAGT 1088
Db 383 GlyAlaAlaAlaGlnArGThrHisArG-----ArgProProGlyCysProArGSer 399
QY 1087 CGGGTTCGTCATCATGAAGTGA---TGGCGCGCGCGCGCGCGACGCTCGGACCAAGCT 1031
Db 400 AlaArGAsnProGlyCysProArGThrTrpArGArGSerGlyAlaGlnArGlyHis 419
QY 1030 CGTACTTGATTTGCGCGCGCTGCGCGGACCTCATCCAGTTCGCGCGCGCTGAGTTGG 971
Db 420 ProProProGlyAlaGlyGlnArGProSerGlyProThrGlyGlyArGProAlaAlaPro 439
QY 970 GGAAGGCTTGCACCAAGTATGATGCTGCGCGCGCGACAGACAGCGGCTGCT--- 914
Db 440 GlyAlaProGlyThr-----ProAlaAlaProGlyProGlyGly 452
QY 913 -----TGAGCATGTTGAGACCGCGACGTCATGGGGATACCGCTCCAGCGAGA 869
Db 453 GlyAlaAlaValProSerGlyAlaThrProHisProGlnArGlySerGlyProAlaAsp 472
QY 868 CGTCCGCGCGCG----- 836
Db 473 ProProAlaAlaAlaArGLeuProProGlyArGlnGlnProArGLeuProGlnAspLeu 492
QY 835 GCAAGCTGTCGCGCGAGCGCGCAATTGGGCTGGCCCATGACCTGTGTGGTGAGATCA 776
Db 493 AlaAlaAlaGlnArG-----CysProAlaGlyPro 502
QY 775 CCACGCGCACGT-----CCAGTCCGCGCGCATGCGGAAGTTCTCTCCAGCGCGGCA 722
Db 503 ProProThrArGSerGlyAlaAlaAlaGlnArGThrHisArGArGProProGlyCysPro 522
QY 721 CGATGTCGCGCTGCTGCGCGCATGCGGTGCGGTAGATGCGCGCATGCGCGGTACAGG 662
Db 523 ArGSerAlaArG-----Asn 527

```

OY	137	GCGACGCAATCCGCCGCATTCGGCATCTGTCTGGAGCCGCCGCCAGCACCCGGGTGCAGC	196
Dd	55	GluaLaValaLAbro-----AlacysHIsalAaplaArgTh--AlaGlyArg--	70
OY	197	TGTTCGGGCGTGGTGCGCGCCCGGTTTCATGACTGCCACACCAGACACAATACC	256
Dd	71	-----GlyArg-----AspProArgHisGlnCysAs	79
OY	257	TGCTCAGCGCGTCCGCATGACGCCCAAGATCTCCAGCGCGTGCACCAACGGT	--GG 310
Dd	79	palaArg-----LeuArgGlyArgHisArgProArgLeuGI	91
OY	311	TCACGGGCAATTCCGG-----	326
Dd	91	yHIsalAvalValArgAspArgLyLeuGIyProHIsGlnArgAspCysValArgLyAl	111
OY	327	-----CATAGCGTCCGCCCGCTGGCGACG	352
Dd	111	aCySHIsGLyProSerValHIsArgAlaAspTrpHisGlnProArgAlaIleuGI	131
OY	353	C-----CAAC	358
Dd	131	uLeuValIGlyInLeuValHIsProGlnSerGlyLeuValProValValArgLyAlPr	151
OY	359	CGCGCGCCCCCTGGACCTGCTGGACGCAAGCGCGCTTACCGTTTGCAGCGCTTGGCG	418
Dd	151	oAlaProProArgParArgParArgParArgPArgBISerArgProArgArgAlaLeuValGI	171
OY	419	ACTACCT---GGACGCGTGGCGGGGCGACCGCGGGCGGCGTCAACGCCCGCTGATGGTG	475
Dd	171	nLeuProLeuProArgParArgParArgGluGlnProGlyGlyHIsGln----	185
OY	476	GCCATTCAAAGCTGGCGCGCGGTCATGCCGGACTTGACGCGCGCGCCACCGAGAG	535
Dd	186	-ProHIsasPaLa-----	189
OY	536	AATATGGCGGCATGCGGGACCTGGCCCGAGAAGCATGCGACGCGCGCCATCGGCATTT	595
Dd	190	AsPaLaValHIsGlyGly----GlyArgGlyValArgGlyGln----	202
OY	596	CGACCGCGCGCTTACCGCGCGCGCGCGCGCGACCAACCGA--AGATCATTCGAG	652
Dd	203	-----LeuGlnSerArgParArgParArgValAlaArgAsHIsArgAs	218
OY	653	TGTGCCGCGC-----GCTGACGCGCGCATG	676
Dd	218	pGIyProSerAspLyArgSerPheAspProProAspArgGlnArgAlaValAlaArgGyl	238
OY	677	GCGGCATTCAGCCAC-----CC	694
Dd	238	paArgLyAlArgTrgHIsArgLeuValAspArgArgValAlaArgProArgGluGlnProValAr	258
OY	695	ACATCGCGAGCAAGS-----CGAGCAATCGTGGCGCGCGCGTGGAGAAACT	742
Dd	258	oaArgValAlaArgGlnArgLyProGlyValArgAlaHIsGly-----	271
OY	743	TCCGCATCGCGCGAGCTGAGCTGCGCGCGGTGTATCTGCACCAACAAGTCATGGCGC	802
Dd	272	-----HIsArgArgArgTrgLyProLeuHIsArgLeuValGlyAlaArgAlaIla--	289
OY	803	AGCCCAATTCGCGCGCGTCCCGGACACSGCTGCGCGCTATCGAGCGCGCCATCGCGCGC	862
Dd	290	-AlaHIsProArgSerValGlnArgProIle-----AspHIsGly-----	303
OY	863	AGGACGCTGCTGCGAGCGGTATCTCTTAGCTGGCGCGCGTCCACCATGTGCTAACAGAGACC	922
Dd	304	-GlyArgGlnLeuHIsArgValGlyAlaGlnIleLyArgLyAlaArgGlyAsaArgGlyAs	323
OY	923	GCGTGCCTGCGCGGAGCGACCATATCATCACTGGTGCAGCGCTTCCCGCACTGACG	982
Dd	323	P---GlyAlaGlyArg-----	327

RESULT 11
S48478
glucan 1,4-alpha-glucosylase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: extracellular glucamylase; mucin-like protein MUC1; protein YK0190
C:Species: *Saccharomyces cerevisiae*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S68478; A26877; B26877; S27281; J06123
R:Rowley R.
submitted to the EMBL Data Library, October 1994

Alignment Scores:	
Pred. No.:	1.06e-08
Score:	296.50
Percent Similarity:	38.72%
Best local Similarity:	25.50%
Query Match:	8.99%
DB:	1
Length:	1367
Matches:	139
Conservative:	72
Mismatches:	266
Indels:	68
Gaps:	13

```

US-10-009-782-1 (1-1758) x s48478 (1-1367)
QY      8 ACTGATGCCGGAGAGAGATTTCACATGTCCCATCGCATCCGACCCTTGACCTGC   67
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     345 ThnSerSerThrThgGlusSerSerAlaProValProThrProSerSerThrThr   36
QY      68 TGCTGCGGGCGGCAACCTCATCGCAGCACACACACCCCGGGGGCGCGGACTG   12
        ||| :||:||||| :||:||||| :||| ||| ||| ||| ||| ||| |||
Db     365 GlusSerSerAlaProValThnSerSerThrThglusSerSerAlaPro-----   38
QY      128 GCGTGGCGGGCAGCCGCAATGCCGCCATGGCGCATGTGTGGAGCGCGCGCGCACCC   18
        ||| :||:||||| :||:||||| :||| ||| ||| ||| ||| ||| |||
Db     383 ValThnSerSerThrThgGlusSerSerAla-----ProValProThrPro   39
QY      188 GGATCGAGATGTGGGCGCTGTGTGTGGCGCGCGGCTTCATGCACCTGCACA-----   23
        |||:||| :||:||||| :||| ||| ||| ||| ||| ||| ||| |||
Db     399 SerSerSerThrThnglUsSerSerAlaProValThnSerSerThrThglusSer   418
QY      239 -----CCCAGCAGCACAATACTGCTGCTCAGGGGTGCAGCATGAGCCGCAAGATCGC   29
        ||| ||| :||:||||| :||| ||| ||| ||| ||| ||| ||| |||
Db     419 SerAlaProValThnSerSerThrThglusSerSerAla-----ProValThnSer   436

```



```

Db      760 SerSerAlaPro---ValProThrProSerSerSerThrThrLubSerSerAlaPro 778
      |||::: ||| ||| ||| |||::: |||
Oy      1391 GCATCCATTCGGTACGTACGTACGCGCCGCCGGTCTGGCAAGACGAGCGGTCACCGGCC 1450
      |||::: ||| ||| ||| |||::: |||
Db      779 ValProThrProSerSerSerSerThrThrLubSerSerSerAlaProValProThrProSer 798
      |||::: ||| ||| ||| |||::: |||
Oy      1451 AGCATGCCGCGCGCGTCTGCTCGACGCACGACGCGCCGCTGAGCCGCGGCGCCAGCTTACAT 1510
      ||| ||| ||| ||| ||| |||::: |||
Db      799 SerSerThrThrLubSerSerValAlaProValProThrProSerSerSer---Asn 817
      |||::: ||| ||| ||| |||::: |||
Oy      1511 CCGGCGTGAACGCGGCGCGCTGCCGCCCTCCCAACCTGAGACGCAACCGCTACATGG 1570
      |||::: ||| ||| ||| |||::: |||
Db      818 IleThrSerSerAlaProSerSerThrThrProSerSerSerThrLubSerSerVal 837
      |||::: ||| ||| ||| |||::: |||
Oy      1571 CCCCCTCCCT 1579
      |||::: |||
Db      838 ProValPro 840

RESULT 12
T02345
hypothetical protein KIAA0324 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02345
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson,
re, J.; White, S.; Deng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: 214664
A:Accession: T02345
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AC08453.1; PID:g2996650
C:Genetics:
A:Map position: 16
A:Introns: 1610/2; 1706/2
A>Note: KIAA0324

Alignment Scores:
Pred. No.: 1,15e-08 Length: 1791
Score: 296.00 Matches: 179
Percent Similarity: 35.61% Conservative: 50
Best Local Similarity: 27.84% Mismatches: 221
Query Match: 9.13% Indels: 194
Gaps: 27

US-10-009-782-1 (1-1758) x T02345 (1-1791)
Oy      1747 AGCTCAGGCGCATCTGGCCCAAGCTCGGCGCTTGACAGAGGTCGAGACG----- 1700
      ||| ||| ||| ||| |||::: ||| |||
Db      619 SerLysSerArgLysSerProArgArgSerArgSerArgSerArgSerProGluValLysAsp 638
      ||| ||| ||| ||| |||::: ||| |||
Oy      1699 CGATCCGCTGCGCTTCGGACAGAGCTTTCGACGAGGCGATTTCCGCGCCCTGCGCTCGCGCC 1640
      ||| ||| ||| ||| ||| |||::: |||
Db      639 LysProArgAlaAlaProArgAlaGlnSerArgSerArgSerArgPro----- 654
      ||| ||| ||| ||| ||| |||::: |||
Oy      1639 GCAGCGCGCGCCCATCATCTCTTGGCCACGATATCGGGTGGGCGCGATATGCGAGCGG 1580
      ||| ||| ||| ||| ||| |||::: |||
Db      655 -----GluProLysAlaProAlaPro-----ArgAlaLeuProArgArgSerArg 669
      ||| ||| ||| ||| ||| |||::: |||
Oy      1579 AGGAGAGGGGCCATGTACGCGTTTGCGTCCAGGGGTGGAGGGGCGCGACG----- 1529
      ||| ||| ||| ||| ||| |||::: |||
Db      670 SerGlySerSerSerLysGlyArgGlyProSerProGluGlySerSerSerThrGluSer 689
      ||| ||| ||| ||| ||| |||::: |||
Oy      1528 -----CCGCCCGCTTCACGCGCGGATGTATAGAGGCTGGGCGCGGCTCAGCGCG 1481
      ||| ||| ||| ||| ||| |||::: |||
Db      690 SerProGluHisProProLysSerArg----- 698
      ||| ||| ||| ||| ||| |||::: |||
Oy      1480 CCGTGCCTGGACACGCGCGCGCATGCTGGCCGCGTGAACCGCTGCTTGGCAGACG 1421
      ||| ||| ||| ||| ||| |||::: |||
Db      699 -----ThrAlaArgArgGlySerArgSerSerPro----- 708
      ||| ||| ||| ||| ||| |||::: |||

```

```

QY 1420 GCGCGCGGTGACGACGAGATGATCGCGCGCGCTGCGTAGGTGTGGAAG 1361
DB -----
QY 1360 TGGCGGTATCGGCCACCGTGGCGGTGACACACACAGTGGCGGAATCGCGCT 1301
DB -----
QY 1300 GCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1241
DB -----
QY 1240 CCGCGGTCTCCAGCGGAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1193
DB -----
QY 1192 GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1133
DB -----
QY 1132 CCGAGCGGTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1079
DB -----
QY 1078 CCATCATAGATAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1019
DB -----
QY 1018 TGGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 959
DB -----
QY 958 ACCAGGTGATGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 899
DB -----
QY 836 -----
DB -----
QY 836 CCGCGCGGTGAGGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
DB -----
QY 837 ArgSerArgSerArgValThrArgArgArgArgArgArgArgArgArgArg 856
DB -----
QY 838 GCGCGCA-----GGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788
DB -----
QY 857 ProAlaArgGlnGlnSerSerArgThrsSerArgArgArgArgArgArgArg 872
DB -----
QY 787 GGTGCGAGATCACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728
DB -----
QY 873 ArgSerArgThrsProThrsProThrsArgThrsArgThrsArgThrsArg 891
DB -----
QY 727 CCGCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 677
DB -----
QY 892 -----
DB -----
QY 676 -----CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 629
DB -----
QY 901 SerProAlaThrHisArgArgSerArgSerArgThrsProLeuLeuSerArg 920
DB -----
QY 628 CCGCGCGCGCGCG-----GCGGATGAGAGCGCGCGCGCGCGCGCGCGCG 578
DB -----
QY 921 ArgSerArgThrsSerProValSerArgArgArgArgArgArgArgArgArg 940
DB -----
QY 577 TGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 518
DB -----
QY 941 -----ArgArgSerArgSer 945
DB -----
QY 517 GGTGCAAGTCCGCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 458
DB -----
QY 946 ArgAlaSerProValSerArgArgArgSerArgThrsProProValThrArg 965
DB -----
QY 457 TGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 422
DB -----
QY 966 ArgSerArgSerArgThrsProThrsArgArgArgSerArgThrsProProVal 985

```

```

QY 421 AGTCGCGCGAGCGCTCGAAGCGGTAAAGCGCGCGCTGCTGCG-----ACGAGTCCA 371
DB -----
QY 966 ThrArgArg-ArgSerArgSerArgThrsProProValThrArgArgArgSerArg 1005
DB -----
QY 370 GGGGGCGCGCGCGGTGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 324
DB -----
QY 1005 rgtThrsProLeuThrArgArgArgSerArgThrsArgThrsProValThrArgArg 1025
DB -----
QY 323 -----CAATGCGCGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278
DB -----
QY 1025 rgsSerArgSerArgThrsSerProVal-----
DB -----
QY 277 TCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 218
DB -----
QY 1035 -----ArgArgArgSerArgSerArgThrsSerProVal-ThrArgArgArgSer 1051
DB -----
QY 217 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 176
DB -----
QY 1052 SerArgThrsProProAlaThrArgArgArgSerArgThrsProLeuLeuProArg 1071
DB -----
QY 175 CCGCGGTGCG-----GACAGATGCGCGCGCGCGCGCGCGCGCGCGCGCG 126
DB -----
QY 1072 LysArgSerArgSerArgSerArgSerProLeuAlaThrArgArgArgSerArgThrsPro 1091
DB -----
QY 125 AGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66
DB -----
QY 1092 ArgThrAlaArgGlnLysArgSerLeu-----ThrArgSerProProAlaThrArg 1108
DB -----
QY 65 AGG-----TCGAGAGCGGTGCGGATGCGGATGCGGATGCGGATGCGGATG 12
DB -----
QY 1109 ArgArgSerArgAlaSerArgSerSerSerArgSerArgSerArgThrsProProAlaThr 1128
DB -----
QY 11 ARG 9
DB 1129 Arg 1129

RESULT 13
D70928
hypothetical protein Rv2913c - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_rev150
C/Accession: D70928
R/Col: S.T.; Broesch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gor,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete g
A/Reference number: A70500; PMID:98295987; PMID:9634230
A/Accession: D70928
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-611 (COL>)
A/Cross-references: GB:274024; GB:AL123456; NID:93250700; PIDN:CA98380.1; PID:el3
C/Genetics:
A/Genes: Rv2913c

Alignment Scores:
Pred. No.: 1,34e-08
Score: 294.00
Percent Similarity: 36.48
Best Local Similarity: 24.06%
Query Match: 8.91%
DB: 2
Gaps: 25

US-10-009-782-1 (1-1758) x D70928 (1-611)
QY 58 TTGACATGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 117
DB -----
QY 16 TysAspValIleIleArgAspGlyLeuThrPheAspGlyThrGlyAsnAlaProLeuThr 35
DB -----
QY 118 GCGGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 171

```

```

Db 36 ArgThrLeuGlyIleArgAspGlyValValAlaThrValAlaAlaGlyAlaLeuAspIu 55
QY 172 GCGCCCGCGCACCGCGGCTGAGCTGTGGCGGCTGTGGCGCGCGCTCATCGAC 231
Db 56 ThrGlyCysProGluValAlaValAlaGlyIleYstrValValProGlyPheIleAsp 75
QY 232 TCGCACACCGCGCGCAACTACTGCTGAGGCGTCCGCGACGACGCGCAAGATCTCG 291
Db 76 ValHisThrHisThrAlaGluValLeuLeuAspProGlyLeuArgGluSerValArg 95
QY 292 CAGGCGCTCACACCGGTGTACGCGCAATTCGCGCATCGCGCGCGCGCTGCGCGAC 351
Db 96 HisGlyValThrThrValLeuLeuGlyAsnGlySerLeuSer-----ThrValTyr 112
QY 352 GCGAACCGCGCGCGCGCGCGCTGAGCTGTGACGACGACG----- 390
Db 113 AlaHisSerLeuAspAlaAlaAspLeuPheSerArgValGluAlaValProArgGluPhe 132
QY 391 -----GCTCTTACCGT-----TTGAGCGCTTCGCGCATCTGCGACGCG 432
Db 133 ValLeuGlyAlaLeuArgAspAsnGlnThrTyrSerThrProAlaGluTyrIleGluAla 152
QY 433 TTCGCGCGCGCGCGCGCGCGCTGACGCGCGCTGTATGTGTGGCGCATCAACGCTGCG 492
Db 153 IleAspAlaLeuProLeuGlyProAsnValSerSerLeuLeuGlyHisSerAspLeuArg 172
QY 493 GCGCGCTGTGCGCGACTGTGACGCGCGCGCG-----ACCGACGAG 534
Db 173 ThrAlaValLeu---GlyLeuAspArgAlaThrAspAspThrValAlaArgProThrGluAla 191
QY 535 GAAATCGCGCGCGCGCGCGCGCTGCGCGAGAGAGCCATGCGCGCGCGCGCGCGCAT 594
Db 192 GluLeuAlaIleMetAlaIleLeuLeuAspGluAlaLeuGlyMetIleGlyMet 211
QY 595 TCGACCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGAGATCAT----- 647
Db 212 Ser-----GlyMetAspAlaAlaIleAspIleAspIleAspGlyAsp 224
QY 648 -----CGAGTGTGCGCGCGCGCGCGCGCATGCGCGCATGAGCGCATACCGCGAC 695
Db 225 ArgPheArgSerArgAlaLeuProSerThrPheAlaThrTyrPheGluArgGlyLeu 244
QY 696 CATGCGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755
Db 245 IleSerValLeuArg-----HisArgGlyArgIle-----LeuGlnSerAlaPro 259
QY 756 CGAGCTGAGCGTGGCGGTGTATCTCGACCGACGAGCATGCGCGCGCGCGCGCG 815
Db 260 -AspValAspAsnProValSerAlaLeuPhePheLeuAlaSerSerArgIlePheAs 279
QY 816 CCGCTCGCGCG-----GAGACGCGTGGCGGT 839
Db 279 nArgArgIleGlyValAlaArgMetSerMetLeuValSerAlaAspAlaIleSerMetPro 299
QY 840 GATC-----GAGCG 848
Db 299 uAlaValHisValPheGlyLeuGlyThrArgValLeuAsnIleLeuGlySerIleVal 319
QY 849 CGCGCATGCGCGCGCGCGCGCGTCTGCGCGCGAT----- 885
Db 319 IArgPheGlnHisLeuProValProPheGluLeuTyrSerAspGlyIleAspLeuProVal 339
QY 886 -----CCCTAGTGGCGCGCGCTGCGCATGCTGCAAG-----CAGACGCG 923
Db 339 IArgGluGluPheGlyAlaGlyThrAlaAlaLeuHisLeuArgAspGluLeuGlnArgAs 359
QY 924 CGTGTGCTGCGCGCGCGCGCGCG----- 945
Db 359 nGluLeuLeuAlaAspArgSerTyrArgArgSerPheArgArgGluPheAspArgIleTyr 379
QY 946 -----ATCATTCACCTGCTGCAAGCG 965

```

```

Db 379 sleuGlyProSerLeuThrPheHisArgAspPheHisAspAlaValIleValGluCys---Pr 398
QY 966 CTTCCCGCACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1025
Db 398 oAspIleSerLeuIleGlyIleYstrSerPheGlyAlaIleAlaAspIleArgIle 415
QY 1026 GTACGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1078
Db 416 -----LeuHisProLeuAspAlaPheLeuAspValLeuValAspAs 429
QY 1079 ACAGAACCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1138
Db 429 nGlyLeuArgAsnValArgTyrThrThrIleValAlaAsnHis-----ArgPhe 445
QY 1139 TCGCGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1195
Db 445 oAsnGluLeuAsnIleValAlaGluProSerValHisMetGlyPheSerAspAlaG 465
QY 1196 GCGACTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1224
Db 465 yAla-HisLeuArgAsnMetAlaPheTyrAsnPheGlyLeuArgLeuLeuValArgAla 485
QY 1225 -----CCGCTGAGACGCGCGGTATGAGAGTGA 1252
Db 485 rAspAlaAspArgAlaGlyGlnProPheLeuSerIleGluArgAlaValTyrArgLeu 505
QY 1253 CCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1312
Db 505 nArgIleLeuAlaGluIleThrPheGlyIle---GlyAlaGlyThrLeuArgGlnGlyAsp 524
QY 1313 TCGCGCGACTGTGTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1372
Db 524 rGluAspPheAlaValIleAspProThrHisLeuAspGluSerValAspGlyTyrHisG 544
QY 1373 CT-----ACCGACGCGCGCGCGCGCG 1393
Db 544 IuGluAlaValProTyrTyrGlyIleuArgArgMetValAsnArgAsnAlaThrVal 564
QY 1394 TCCATTCGCGTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1452
Db 564 aValAlaThrGlyValGlyGlyThrValValPheArgGlyGlnPheGlyGlnP 584
QY 1453 -----CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1486
Db 584 nArgAspIleTyrGlyGlnAsnValIleSerGlyArgTyr----- 597
QY 1487 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1530
Db 598 -----LeuArgAlaGlyGluLeuGlyAlaAla 606

```

RESULT 14
 A:3932
 mucin 2 precursor, intestinal - human (fragments)
 N:Alternate names: mucin SMUC-41
 C:Species: Homo sapiens (man)
 C:Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
 K:Accession: A49963; A45106; B45106; B3532; A61257; P00328; P00329
 R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
 J. Biol. Chem. 269, 2440-2446, 1994
 A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of
 A:Reference number: A49963; MUID:94132002; PMID:8300571
 A:Accession: A49963
 A:Molecule type: mRNA
 A:Residues: 1-639 <GU>
 A:Cross-references: GB:L21998
 J. Biol. Chem. 267, 21375-21383, 1992
 A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
 A:Reference number: A45106; MUID:93016075; PMID:1400449
 A:Accession: A45106
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 626-1895 <GU>

```

Db 1403 ThrThrThrProSerProProProThrThrThrThrLeuPro-----Pro 1411B
QY 122 ACCTGGGCGTGGCGGGACCGCATCGCGGCATCGGCATCTGTGGACGCCGGCGGC 181
Db 1419 Thr-----ThrThrProSerProProThrThrThrThrProProPro 1430
QY 182 ACACCGGCTGACGCTGCGGGCTGGTGGTCGGCGCGGCTGATCATGCTGCACACC 241
Db 1435 ThrThrPro-----SerProProLeThrThrThrThrPro 1448
QY 242 ACGAGCAACTACTGCTCAGGCGTCGGGACATGACGCCAAGATCTCGAGGGCGCA 301
Db 1449 LeuProThrThr-----ProSer 1455
QY 302 CCAAGGGGTGACGGGCAATTGGGATCAGCTGGCGCGGTGGCGACGCCACCGC 361
Db 1456 ProProLeSer-----ThrThrThrThr 1463
QY 362 CCGCGCCCGCTGGACCTGCTGGACGAGAGCGGCTCTTACCGTTTGAGCGGTTCGCGCAGT 421
Db 1464 ProProProThr-----ThrProProPro 1473
QY 422 ACCTGGACCGCTTGGCGGCACGCGGCGCGCTGACAGCGCGCTGTATGTGGCCATT 481
Db 1474 ThrThrThrProSerProProThrThrThrProSerProProThr-----1489
QY 482 CAACGCTGGCGCGCGGCTATGGCGGACTGTGACGCGCGCGCGACGAGAGAAATG 541
Db 1490-----ThrThrThrProProProThrThrThrPro 1500
QY 542 CGGCGATCGGGACCTGGCGCGAGAGACCAATGGCGAGCGGCGCATCGGCATTGGACCG 601
Db 1501 SerProProMetThrThrThrPro-----Pro 1511
QY 602 GCGCGTTGTACCGCGCGCGCGCGCGCGCCACCAACGAGAAATCATGAGTGTGCGCGC 661
Db 1512 AlaSerThrThrThrLeuProProThrThrThrPro-----1523
QY 662 CGGTGACGGCGATGGCGGCGATCATGCCACCAACATGGCGACGAGCGACATCG 721
Db 1524-----SerProProThrThrThrThr-----ThrThr 1532
QY 722 TGGCGGCGTGGAGAAACCTTCGCGATGGCGCGAGCTGGAGTACCT--ACGTGGCG 898
Db 1533 ProProProThrThrThrProSerProProThr-----1543
QY 782 CGCACCAAGGTCATGGGCGCGACGCCAATTTCGCGCGCTCGCGGAGACGTCGCGCTGA 841
Db 1544-----ThrThrProLeThr-----1548
QY 842 TCGAGCGCGCATGGCGCGCGCAGGACGTCGTGTCGAGCGGTACCT--ACGTGGCG 898
Db 1549-----ProPro-----ThrSerThrThrLeuProProThrThrPro 1562
QY 899 GCTCGACCATGCTCAGGAGAGACCGCGTGTGCTGCGCGGACGACCATCATCCTGCT 958
Db 1563 SerProProProThrThrThrThr-----ProProProThr 1575
QY 959 GCAACCCCTCCGGAACCTGAGCGGCGACCTGATGAATCGGGCGGACCGGCA 1018
Db 1576 ThrThrProSerPro-----ProThrThrThr 1584
QY 1019 AATCAAGT-----ACGAGTGTGCGCGAGTGCAGCGCGCGCGCGCA 1063
Db 1585 ThrProSerProProThrThrThrThrThrProProProThrThrThrProSerPro 1604
QY 1064 TCTACTCTAGATGAGAGAAC-----CGAGGTGAGGAGCATCTCGCGCTTCGCGCGCA 1117
Db 1605 ProThrThrThrThrThrThrProProProProThrThrThrProSerProProThrThr 1624
QY 1118 CCAATGATCGCTCGGAGCGCTCGCGCAGACGAGACGCGCCATCGCGCTGTGGGGA 1177
Db 1625 ProLeThrProProThrSer-----ThrThr-----1733

```

```

QY 1178 CATTCCCGGGTGTGGGCACTATGCGGCGACCTGGGCGCTTCCCGCTGGAGACG 1237
DB 1634 -----ThrlProProThrThr----- 1640
QY 1238 CGGTATGGAAGATACCGCGCTGACCGCGCGCTGCGCGCGGCGGAGC 1297
DB 1641 -----ProSerProProThrThrThrPro----- 1651
QY 1298 TCGAGCGCGGTACTTCGGGACCTGTGTTCACCGCGCGCGAGTGGCGGATA-- 1354
DB 1652 ---ProProThrThrProSerProProThrThrThrProSerProProThr 1670
QY 1355 -----CCGCGACCTTCGAACACCTACCGAGCGCGCG-- 1387
DB 1671 ThrThrThrProProProThrThrThrProSerProThrThrThrProSer 1690
QY 1388 -----CCGCGACCTTCGAATCCGCTG 1405
DB 1691 ProProThrThrThrMetThrThrProSerProThrThrThrProSer---Pro 1709
QY 1406 AGCTACGCGCGCGCGCTGCAAGACGCGCTTCACCGCGCGACATGCCGCGCG 1465
DB 1710 ThrThrThrThrThrProSer-----SerThrThrThrProSerProProThr 1727
QY 1466 TCGTCGACGCGACGCGCGCTGAGCGCGCGCGCGCTTACATCCGCGCGTGAACGG 1525
DB 1728 MetThr-----ThrProSerProThrThrThrProSerProProThrThrMet 1744
QY 1526 CGCGCTGCGCGCGCTCCCAACCTGAGCGGACGCGCGTACATGCGCGCTCCGCTC 1585
DB 1745 ThrThrThrProProThrThrThrSerSerProLeuThrThrThrProLeuPro 1763
QY 1586 CCAATACGCGCGCGCGATATCG 1609
DB 1764 SerLeuThrProProThrPheser 1771

```

RESULT 15

```

JQ0405
hypothetical 119.5K protein (uvra region) - Micrococcus luteus
N:Alternate names: ORF 1 protein
C:Species: Micrococcus luteus, Micrococcus lysodeikticus
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
R:Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A:Title: Micrococcus luteus homolog of the Escherichia coli uvra gene: identification of
A:Reference number: S04781; PMID:89364717; PMID:2549377
A:Accession: JQ0405
A:Molecule type: DNA
A:Residues: 1-1106 <SH2>
A:Cross-references: EMBL:X15867
A>Note: this reading frame extends between two stop codons and does not begin with a start
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

```

Alignment Scores:

```

Pred. No.: 3,18e-08 Length: 1106
Score: 287.50 Matches: 196
Percent Similarity: 32.81% Conservative: 35
Best Local Similarity: 27.84% Mismatches: 243
Query Match: 8.87% Indels: 230
Gaps: 37

```

US-10-009-782-1 (1-1758) x JQ0405 (1-1106)

```

QY 1744 TCAGGCGCATTCGGCGCGCTTGGACAGGCTCGAGACCGCGATCCGCTGGCCT 1685
DB 391 AAlAGlGlylProolProAlrgrg-----ArglInrgrgrgAlaValPro 405
QY 1684 GCGACAGGT-----CTGACAGG 1667
DB 406 AAlArgrglYArgrArgrAlaValAlaAlaArgrGlnValHlslenglyLeuAlaAla 425

```

```

QY 1666 TCAATTCGGGGCTTGGCGTCGCGCGAGCGCGCGCCATCACTTCCTTCCACAGA 1607
DB 426 Arg-----AlaArgrAlaGlnGlnGlnGlnYArgrglYLeu-----LeuProGly 439
QY 1606 TATCGGTGGGGCGGTATTCGAGCGGAGGAGGCGCGCTGAGCGGT----- 1559
DB 440 HlslProVal-----GlnGlyProAlrgrglYAlaArgr 452
QY 1558 -----TTCGCTCGAGGCTGGAGGCGCGCG-----ACGCGCGCGCTTCACGC 1514
DB 453 ArgrProAlArgrGlnGlnGlnGlnGlnGlnYArgrValProGlnProValArgr 472
QY 1513 CGAATGTAGAGGCTGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCG-- 1457
DB 473 AlArProLeuHlslGlnGlnGlnGlnGlnGlnYArgrHlslProLeuArgrAlaGlnAla 492
QY 1456 -----CATGCTGCGCGGTGAACGCT 1436
DB 493 AsparGValGlyArgrArgrProArgrAlaLeuArgrValAlaHlslAlaGlnYArgrProValPro 512
QY 1435 GCTCTGCGAGACCGCGCGCGCGCGCTGACGTACACGGAATGGATCGCGCGCGCTCG 1376
DB 513 GlnYAlProArgrGlyProProGlnProHlsl----- 522
QY 1375 TAGGCTTTCGAGAGTGGCGCGCTATCGCGCGCGCGCGCGCGCTGCAACACCA-- 1322
DB 523 ---GlyProGlnArgrAlaArgrGlnYArgrProValHlslArgrGlnHlslProAlaHlsl 541
QY 1321 -----GTCGCGGAATACCGCGCGCGCTGACGTGCGCGCGCGCGCGCG 1286
DB 542 AlArgrGlnHlslGlnYAlaLeuLeuGlnYAlaAlaAlaArgrglYProGlnYAlaAlaArgr 561
QY 1285 CGCGCGCGCGAGACCGCGCGCGCGCGCGCGCGCGCGCTTCATCCGCGCGCTCGACGC 1226
DB 562 GlnYProGlnAlaGlnGln-----AspProGlnProAla 572
QY 1225 GGAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1172
DB 573 GlnYAlProAlaGlnYArgrArgrProArgrValProGlnProArgrAlaGlnYArgrHlsl 592
QY 1171 ACAGCGCGCGAGTGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1112
DB 593 Leu---ArgrArgrGlnGlnProAlaHlslProProGlnHlslThrsArgrglYArgrAlaGly 611
QY 1111 CGAAGCGCGAGATGCGCGTGCAGTGGGT-----CGTCCATCAATGAATGAATGCGCGCG 1055
DB 612 ArgrArgrProLeuArgrProArgrAlaValAlaHlslArgrProAla----- 625
QY 1054 CGCGCTGCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAT 995
DB 626 ProAlaGlnYAlaProProProHlslArgrArgrProProAlaProAlaGlnYArgrHlsl 645
QY 994 CCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG--AGGCGTTCG 959
DB 646 ProHlslArgrAlaArgrAlaArgrGlnHlslArgrArgrGlnYArgrAlaArgrglY----- 664
QY 958 ACCAGGTATGATGGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 899
DB 665 -----HlslArgrPro---SerArgrGly----- 670
QY 898 CGCGCAGTAGGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-- 857
DB 671 ArgrValArgrArgrGlnYArgrAlaGlnGlnGlnGlnGlnYArgrProGlnYArgrGlnGlnHlslAla 690
QY 856 -----CCAGCGCGCGCT----- 845
DB 691 ValArgrHlslArgrArgrLeuArgrProArgrProLeuHlslArgrGlnYAlaAlaSer 710
QY 844 CGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGGT 785
DB 711 ArgrProGlnYArgrAlaArgrAlaArgrglYProArgrArgrProGlnYArgrGlnGlnGln 730

```

```
OY 784 GCGAGATACCA---CCGGCAGCTCCAGCTCGCGCCGATCGGAGATTTCCTCCAGCG 728
DB 731 ArgLeuGlyProGlyProAlaArgGlyProHisGlyArgAspGlyArg----- 746
OY 727 CCGCCACGATGCTGCTCGCTTCGTCGCGCATGTGGGTGGCTAGA----- 683
DB 747 -----ValArgLeuArgGlnValHisAlaAspGlnArgAspProLeuGlnGly 762
OY 682 -----TGGCGCATGGCGCTCAGCGCGCGCCGACACTCGATGATCT 641
DB 763 ProGlyGlnProAlaGlnArgArgGlnAlaArgAla-ArgProAlaProVal----- 779
OY 640 CTTCGGTGGTGGCGCGCGCGCGGTAGAGCGCGCT-----CGAATGCCGA 587
DB 780 -----ArgGlyGlyThrArgAlaProGlyGlnGlyProAlaGlyProGly 795
OY 586 TGGCGCGCTGGCCATGGCTCTC-----GGCCAGTCT----- 553
DB 795 uProHisArgAlaHisAlaThrLeuGlnProArgHisLeuHisGlyArgValArgArgAs 815
OY 552 -----CCGATGGCGCGCATTTCTCGTCGGTGGCGCGCGCT 515
DB 815 PProGlnAlaLeuArgGlyAspProArgGlyGlnGlyProGlyLeuProAlaGlyProVa 835
OY 514 GCAATCCCGCATGACCGCGCGCGCGCATGAAATGGCCACCATACAGCGCGCTGA 455
DB 835 lLeuLeuGlnHisGlnGlyArgAlaLeuArg-----GlyValArg 848
OY 454 CCGCGCGCGCGCTGGCGCGCGCATCGCTCCAGGTAGTCCGCGAAGCCTCGAAGCGTAG 395
DB 848 gGlyArgArgHisAlaGlnAspArgArgPLeuProAlaGlyArgLeuArgAlaValArg 868
OY 394 AGCGCGCTTCCTCCAGCAGCTCCAGGGGCGC-----GGCGCGGTGGCGTGGCGCA 344
DB 868 gGlyValProArgGlyProValGlnProGlyAspAlaArgGlyHisLeuGlnGlyGlnG 888
OY 343 GCGCGCGCGCATGATCGCGCATGTCGCTGACCGCTGTGACGCGCTCGAGATCT 284
DB 888 uHisArgArgGlyProArgHisAlaAspArgGlyArgGly---LeuLeuGlnArgGly 907
OY 283 TGGCGCTCATGCTCGCAGCGCTGAGCAGGTGTGCTGCTGGGTGGCGATGATGA 224
DB 907 lHisProHisLeuAlaValPro--GlyHisAlaArgArgArg----- 921
OY 223 AGCGCGCGCGCAGCAGCGCGCAGCAGCTGACCGCTGTGCGCGCGCGCTCGACA 164
DB 922 SerGlyLeuArgProSerGlyProAlaArgHisHisAlaLeuGlyArgArgGlyProAla 941
OY 163 GATCGCGATGGCGCGCATGGCTCGCGCA-----CGCCCA----- 125
DB 942 -----ArgGlnAlaGlyGlyArgAlaProGlnAlaLeuGlnArgProHisHis 957
OY 124 -----GTCGCGCGCGCGCGCGCGGTGTCGCGCGCATGAGGTCGCGCGCGCA 71
DB 958 LeuArgAlaGlyArgAlaAspHisGlyValAlaLeuArgArg-----HisProGln 974
OY 70 GCAAGCAGGTGGAAGGCTGGGAATCGGATGGACATGGAATCTCTCCGCGCATCA 11
DB 975 AlaProAlaArgAlaSer-----ValProAlaGlyGly 984
OY 10 AGTGA 5
DB 985 GlnGly 986
```

Search completed: May 11, 2003, 12:12:51
Job time : 84.5 secs

GenCore version 5.1.4-P5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_nzp model

Run on: May 11, 2003, 12:12:55 ; Search time 77 Seconds
(without alignments)
7808.619 Million cell updates/sec

Title: US-10-009-782-1
Sequence: 1 gaattcactgacgcgga.....ccctgagctacgagaagctt 1758

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 787003 seqs, 171007862 residues

Total number of hits satisfying chosen parameters: 1574006

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+np.model -DEV=xlp
-O/cgcn2.1/USPTO.spool/US10009782/runat_07052003_122519_23229/app_query.fasta.1.1927
-DB=Pending Patents_AA.New -QWMT=blastn -SUFFIX=rapn -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl
-LIST=45 -DOCALL=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTWMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10009782.ccgcn.1.1.42.gunat.07052003_122519_23229 -MCPU=6 -ICPU=3
-NO_XLPEXT -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-NARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending Patents_AA.New:*
1: /cgcn2.6/ptodata/2/paa/US06_NEW_COMB.pep.*
2: /cgcn2.6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgcn2.6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgcn2.6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgcn2.6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgcn2.6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgcn2.6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1111.5	33.7	475	6	US-10-282-122A-50211
2	1106.5	33.5	479	6	US-10-282-122A-60214
3	659	20.0	530	6	US-10-156-761-10819
4	539	16.3	576	6	US-10-156-761-11333
5	455	13.8	2294	6	US-10-366-683-17231
6	455	13.8	2294	6	US-10-366-683-17231
7	434	13.2	885	6	US-10-419-128-17231
8	434	13.2	885	6	US-10-366-683-26129
9	433	13.1	937	6	US-10-419-128-26129
10	433	13.1	937	6	US-10-366-683-19446
11	432.5	13.1	663	6	US-10-419-128-19446
					Sequence 30843, A

12	432.5	13.1	663	6	US-10-419-128-30843	Sequence 30843, A
13	431.5	13.1	822	6	US-10-366-683-21920	Sequence 21920, A
14	431.5	13.1	822	6	US-10-419-128-21920	Sequence 21920, A
15	430.5	13.0	651	6	US-10-366-683-32204	Sequence 32204, A
16	430.5	13.0	651	6	US-10-419-128-32204	Sequence 32204, A
17	424.5	12.9	1002	6	US-10-366-683-27980	Sequence 27980, A
18	424.5	12.9	1149	6	US-10-419-128-27980	Sequence 27980, A
19	424.5	12.9	1149	6	US-10-366-683-25557	Sequence 25557, A
20	424.5	12.9	1149	6	US-10-419-128-25557	Sequence 25557, A
21	423	12.8	801	6	US-10-366-683-29274	Sequence 29274, A
22	423	12.8	801	6	US-10-419-128-29274	Sequence 29274, A
23	421.5	12.8	863	6	US-10-366-683-26099	Sequence 26099, A
24	421.5	12.8	863	6	US-10-419-128-26099	Sequence 26099, A
25	413.5	12.5	774	6	US-10-366-683-16789	Sequence 16789, A
26	413.5	12.5	774	6	US-10-419-128-16789	Sequence 16789, A
27	410	12.4	1706	6	US-10-366-683-31760	Sequence 31760, A
28	410	12.4	1706	6	US-10-419-128-31760	Sequence 31760, A
29	408.5	12.4	638	6	US-10-366-683-27068	Sequence 27068, A
30	408.5	12.4	638	6	US-10-419-128-27068	Sequence 27068, A
31	407.5	12.4	672	6	US-10-366-683-16941	Sequence 16941, A
32	407.5	12.4	672	6	US-10-419-128-16941	Sequence 16941, A
33	407	12.3	639	6	US-10-366-683-20571	Sequence 20571, A
34	407	12.3	639	6	US-10-419-128-20571	Sequence 20571, A
35	404	12.2	1053	6	US-10-366-683-26140	Sequence 26140, A
36	404	12.2	1053	6	US-10-419-128-26140	Sequence 26140, A
37	402.5	12.2	1418	6	US-10-366-683-32367	Sequence 32367, A
38	402.5	12.2	1418	6	US-10-419-128-32367	Sequence 32367, A
39	401.5	12.2	639	6	US-10-366-683-28453	Sequence 28453, A
40	401.5	12.2	639	6	US-10-419-128-28453	Sequence 28453, A
41	399	12.1	724	6	US-10-366-683-31715	Sequence 31715, A
42	399	12.1	724	6	US-10-419-128-31715	Sequence 31715, A
43	395	12.0	802	6	US-10-366-683-25050	Sequence 25050, A
44	395	12.0	802	6	US-10-419-128-25050	Sequence 25050, A
45	394	11.9	631	6	US-10-366-683-30909	Sequence 30909, A

ALIGNMENTS

RESULT 1
US-10-282-122A-50211
Sequence 50211, Application US/10282122A
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELTRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931

RESULT 2
 US-10-282-122A-60254
 : Sequence 60254, Application US/10282122A
 : GENERAL INFORMATION:
 : APPLICANT: Wang, Liangsu
 : APPLICANT: Zamudio, Carlos
 : APPLICANT: Malone, Cheryl
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Karl
 : APPLICANT: Zyskind, Judith
 : APPLICANT: Wall, Daniel
 : APPLICANT: Trawick, John
 : APPLICANT: Carr, Grant
 : APPLICANT: Yamamoto, Robert
 : APPLICANT: Forsyth, R.
 : APPLICANT: Xu, H.
 : TIME OF INVENTION: Identification of Essential Genes in Microorganisms
 : FILE REFERENCE: ELITRA.034A
 : CURRENT APPLICATION NUMBER: US/10/282,122A
 : CURRENT FILING DATE: 2003-02-20
 : PRIOR APPLICATION NUMBER: 60/151,078
 : PRIOR FILING DATE: 2000-03-21

CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 10819
 LENGTH: 530
 TYPE: PRT
 ORGANISM: Streptomyces avermitilis
 US-10-156-761-10819

Alignment Scores:

Pred. No.:	9e-16	Length:	530
Score:	659.00	Matches:	182
Percent Similarity:	46.34%	Conservative:	71
Best Local Similarity:	33.33%	Mismatches:	201
Query Match:	19.98%	Indels:	92
DB:	6	Gaps:	12

US-10-009-782-1 (1-1758) x US-10-156-761-10819 (1-530)

QY 61 GACCTGCTGCTGCGGGGCGACCTCATGCGAGCGACACACCGCGGGCGCGCC 120
 DB 2 AspleuValIleArgAspValAlaArgValAlaAspGlyThrGlyAlaSerCysArgAla 21
 QY 121 GACGTGGCGGGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 DB 22 AspValGlyIleThrGlyAlaArgGlyAlaIleArgValAlaArgValAlaIleGly 41
 QY 181 CACACCGCGGTC--GACGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 237
 DB 42 GlyArgValAlaIleAspAlaGlyGlyLeuAlaLeuAlaProGlyPheIleAspMetHis 61
 QY 238 ACCGAGCGAGCACTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 297
 DB 62 AlaHisSerAspLeuAlaLeuAlaArgAspProAspHisSerAlaLysAlaIleGly 81
 QY 298 GTCAACGAGGTGTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 345
 DB 82 ValThrLeuGluValAlaGlyGlnAspGlyLeuSerGlyAlaProValAspAspArgThr 101
 QY 346 -----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 369
 DB 102 LeuAlaGluValAlaArgAlaIleThrGlyTyrAsnGlyHisGly----- 116
 QY 370 CTGAGCTGCTGCGAGCGAGCGGCTGACCGTTGAGCGGCTTCCGCGACTGAC 429
 DB 117 ---AspAspIleAsp-----PheThrTyrArgSerValGlyGlyTyrLeuAsp 131
 QY 430 GCGTTGGGGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 489
 DB 132 ArgLeu---AspGlnGlyIleAlaValAlaAspAlaIleTyrLeuIleProGlnIleThrVal 150
 QY 490 CGGCGCGGCGTATGCGGAGCTTGCAGCGCGCGCGCGCGCGAGCGAGAAATCCGCGCGATG 549
 DB 151 ArgMetLeuAlaValAlaGlyTyrGluAspArgAlaIleThrProGlyGluLeuAspArgMet 170
 QY 550 CGGAGCTGCGGAGGAGCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 609
 DB 171 ArgArgLeuValAlaAspGlyMetArgGlyGlnAlaValAlaGlyMetSerSerGlyLeuThr 190
 QY 610 TACCGCGCGCGCGCGCGCGCGCGCGAGCGAGATCATGAGGTGCGGCGGCGGCGGCGGCGG 669
 DB 191 TyrThrProGlyMetTyrAlaGluAspAlaGluLeuThrGluLeuCysArgValAla 210
 QY 670 GCGCATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 729
 DB 211 GluTyrGlyGlyTyrTyrCysProHisIleArgSerTyrGlyAlaGlyAlaIleGlnAla 230
 QY 730 CTGAGGAAACCTTCGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 789

DB 231 TyrGluGluMetValAlaLeuThrArgGluAlaArgCysProLeuHisLeuAlaHisAla 250
 QY 790 AAGTCATGGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 849
 DB 251 ThrMetAspPheGlyValAlaAspGlyGlyArgAlaProAspLeuAlaLeuAspGly 270
 QY 850 GCGATGGCGGCGG---CAGGAGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 906
 DB 271 AlaLeuAlaGluGlyAlaAspIleThrLeuAspThrTyrProGlyTyrProGlyCysThr 290
 QY 907 ATGCTCAAGCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 963
 DB 291 ThrLeuVal-----AlaMetLeuProSerTyrAlaGlyGlu 302
 QY 964 CCTTCCCGGAA--CTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
 DB 303 GlyIleProGluAlaValAlaLeuAlaArgLeuAlaAspArgThrValAlaGluThrGly 322
 QY 1021 TCCAAATACGAGCGGCGG--- 1038
 DB 323 HisAlaMetGluValAlaGlyAlaAspGlyCysHisGlyAlaProIleGluTyrAspThr 342
 QY 1039 -----CCGAGCTGCGAGCGG----- 1053
 DB 343 IleGluIleSerGlyAlaSerAspProAlaLeuAlaProTyrValGlySerThrValGln 362
 QY 1054 -----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 DB 363 AlaSerAlaAspLeuArgGlyGluAlaProThrThrAlaArgArgLeuLeuAsp 382
 QY 1081 -----GAAACCGAGCTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1101
 DB 383 AspArgLeuGlyCysThrIleLeuGlnHisValGlyHisGluGlnHisValAlaArgAlaIle 402
 QY 1102 CTGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1161
 DB 403 MetArgHisArgValHisThrGlySerAspArgIleLeuGlnGlyThrIleAspProHis 422
 QY 1162 CCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1221
 DB 423 ProAlaGlyAlaTyrGlyThrPheProHisTyrLeuGlnHisTyrValAlaGluLeuGlyAla 442
 QY 1222 TTCGCGGTGAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1281
 DB 443 LeuSerLeuGluGluCysValAlaHisLeuThrGlyArgProAlaAlaArgLeuGlu 462
 QY 1282 GCGG 1341
 DB 463 ProAspArgGlyLeuValAlaGlyGluTyrArgAlaAspLeuValLeuPheAspProSer 482
 QY 1342 ACGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1401
 DB 483 ThrValAlaAlaGlySerThrPheGluAspProArgAlaGlyLeuProHis 502
 QY 1402 GTGTAAGTCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1461
 DB 503 ValLeuIleAspGlyArgPheValAlaGluAspGlyArgThrArgThrAspValLeuAlaGly 522
 QY 1462 CGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1479
 DB 523 ArgAlaValAlaArgArgThr 528

RESULT 4.
 US-10-156-761-11333
 Sequence 11333, Application US/10156761

GENERAL INFORMATION:
 APPLICANT: OMIURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA

RESULT 5

US-10-366-683-17231
 ; Sequence 17231, Application US/10366683
 ; GENERAL INFORMATION:
 ; APPLICANT: Rubenstein, Marc J.
 ; APPLICANT: Nolling, Jork
 ; APPLICANT: Deloughery, Craig
 ; APPLICANT: Bush, David
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: PAT03-04
 ; CURRENT APPLICATION NUMBER: US/10/366,683
 ; PRIOR FILING DATE: 2003-02-13
 ; PRIOR APPLICATION NUMBER: 09/252,991
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 17231
 ; LENGTH: 2294
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-10-366-683-17231

Alignment Scores:
 Pred. No.: 8,45e-09
 Score: 455.00
 Percent Similarity: 34.51%
 Best Local Similarity: 29.97%
 Query Match: 13.79%
 Length: 2294
 Matches: 238
 Conservative: 36
 Mismatches: 239
 Indels: 281
 Gaps: 46

US-10-009-782-1 (1-1758) x US-10-366-683-17231 (1-2294)

QY 9 CTGATCGCGAGAGAGATTTCATGTCATCCGATCCGACGCTTCGACGCT 68
 DB 675 LeuAlaArgProGlnArgAspAlaArgProProheAlaLeu-----ProAla 690
 QY 69 GCTGCGGCGGCGACCTCATCGACGACGACGACGCGGCGGCGCGGCGACCT 125
 DB 691 AlaValGlyArgTyrAlaGlnHisArgProGlyLeuAlaValAlaValAlaArgProAla 710
 QY 126 -----GGCGCTGCGCGCGA-----CCGCATCGCGCGCATCGGCATCT 164
 DB 711 AspGluAlaGlyArgArgAlaGlyValGlnAlaHisProAlaGlnHisArgAlaArg 730
 QY 165 GTGCGAGCG----- 173
 DB 731 ProGlyAlaGlyHisArgProProGlyGlyAlaAlaAlaGlyAspArgAlaAlaThrPro 750
 QY 174 -----CGCGCGCGACACCGGCGCGACGCTGCGCGCTGCGCGCGCG 221
 DB 751 GlyAlaLeuProArgArgArgGlnGlnGlyProAlaGlyGlyGlyArgArgArg 770
 QY 222 CTTCATCGACTCGACACGACGACGACGACGACGCTGCTCAGGCGCTGCGGACATGACGCC 281
 DB 771 ProLeuGlyGlyAspHisProArgArgAlaValAlaGlnArgGlnArgGlnArgAla 790
 QY 282 CAGATCTGCGAGGCGGTACACAGGTGCTACGCGGCAATTGCGGCATCGCTGCGCGC 341
 DB 791 -----ArgArgValProAlaAla 796
 QY 342 GCTGCGCA-----CGCCACCGCGCGCGCGCTGAGCT-----GCTGAGAGAGCGG 392
 DB 797 AlaAlaAlaHisArgArgArgProAlaGlnGlnGlyProGlyAlaGlyHisArg--A 816
 QY 393 CTTCACGCTTCGAGCGCTTCGCGACCTGACGCGGCTGCGCGCGCGCGCGCGCG 452
 DB 816 lAlaLeuThrThrLeuArgSerPro-----ArgGlyAlaThrAlaThrProThr 832
 QY 453 GGTCAACGCGCGCTGATGTTGGCGCATCAAGCTGGCGCGCGCGCGCGCGCGACTT 512
 DB 832 lYAsnCysPro-----ArgProAlaProArgArgSerProThr 845
 QY 513 GAGGCGCGCGCGCGCGAGAGAAATGCGCGCATGCGGCGCGCGCGCGAGGAGCAT 572

DB 845 eSerAlaArgProPro-----AlaThrProAlaAlaSerProProlys-----G 860
 QY 573 GCCACGCGCGCGCATTCGACGCGCGCGCTTCACCGCGCGCGCGCGCGCGCAC 632
 DB 860 lYAlaAlaThrProSerArgTyrArgProMetProAla-----ProlysAlaAla- 876
 QY 633 CACCGAGAGATCATCGAGGTGTCCGCGCGCGCTGACGCGCGCGCGCGCATTAAGCCAC 692
 DB 877 His-----AlaThrAlaGlyTrp-----LysSer 884
 QY 693 CCACATCGCGCGACGAGCGCGACATCGT-----GGCGCG 728
 DB 885 ProTyrSerArgArgAlaArgAlaArgAlaArgLeuProGlnLeuLeuArgAlaLeuGly 904
 QY 729 GCTGAGGAGAACTTCGCGCATCG-----CCGCGAGCTGAGCGT 767
 DB 905 AlaGlyAlaArg--ProAlaArgProGlnProAlaSerHisLeuValProArgProAlaAla 923
 QY 768 GCGGCTGTGATCTCGCACCAAGATCATGCGCGCGCGCGCGCGCGCGCGCA 827
 DB 924 AlaProGlyGlnLeuArgTyrThrValGlyLeuGlyAsp-----GlnProLeuGlyAla 940
 QY 828 GAC-----GCTGCC 836
 DB 941 AspArgProAlaValProGlyLeuAspArgLeuProArgAlaAlaHisArgAlaGlyPro 960
 QY 837 GCTGATGAGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 896
 DB 961 AlaGlnArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 980
 QY 897 CGGCTCCACCATGCTACGCA-----GGACCG 923
 DB 981 ArgArgHisArgArgArgArgAlaGlyAspProGlnAlaThrHisGlyArgGlyProGlyPro 1000
 QY 924 CGGCTGCG 971
 DB 1001 AlaGlnAlaGlyGlnAlaGlyArgGlnArgAlaAlaGlnProValArgValAlaLeuVal 1020
 QY 972 CGACATGACCG----- 983
 DB 1021 ArgAspHisArgProAlaArgPheGlyGlnAspHisArgAlaAspGluLeuArgProGly 1040
 QY 984 -----GCGGACCTGATGATGATGCGCGCGCGCGCGCGCGCGCG 1016
 DB 1041 LeuProAlaArgArgThrAspGlyArgArgGlyAspProArgGlyArgArgHisAlaGln 1060
 QY 1017 ---CAATTCAGTACGAGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1073
 DB 1061 LeuArgLeuValAlaHisArgArgSerGlyAlaAlaGlyHisArgArgProLeuTyrHis 1080
 QY 1074 GATGACGA-----ACCGGACGTGACCG 1097
 DB 1081 ProGlyGlnProArgProGlyGlyGlnGlyArgLeuAlaGlyLeuProArgProAlaGln 1100
 QY 1098 CATGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119
 DB 1101 AspProAlaGlnAlaProProHis--ArgTyrArgValHisArgHisGlnProPheArg 1119
 QY 1137 CTTGCGCGA-----CGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1169
 DB 1120 ProAlaAlaGlyGlnArgArgArgAlaArgArgProCysProGlyAspProArgThrHis 1139
 QY 1170 GTGGGCGACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1208
 DB 1140 ProGlyAlaValProAlaAlaArgGlyAlaLeuProAspLeuArgAspAlaAspGlnVal 1159
 QY 1209 CGACCTGCG-----CCTGTTCCGCTGAGAC--GCG 1238
 DB 1160 ArgProGlyAlaGlyLeuHisGlyValLeuArgGlnProGlnProArgGlyThrArgAla 1179
 QY 1239 GGTATGAGATGAC--CGGCTGACCGCGCG----- 1268
 DB 1180 GlyLeuGlyAspAspLeuArgProArgArgGlnGlnArgGlyArgProAlaGlyGly 1199

[illegible]

```

Db 1310 ArgAlaAlaValProAlaGlySerGlyValAlaGlyThrArgProGlyIleHisIleProGlu 1329
QY 1569 GGC-----CCCTCCCTCCGCTCGCAATACGAGGCCCAACCCGATATGCT-----GGG 1613
Db 1330 GlyArgAlaProProGluValAlaAspHisAlaGlyAlaGluAlaGlyHisAlaGlyIle 1349
QY 1614 CAAGAGAGTAGTATGGCGCGCGCGCTGCCGCCCGCAGCGCAAGCCCGGAA-----1661
Db 1350 AlaGlyIleAspGlyLeuMetAspArgGlnLeuProArgGlnProGluLeuHisArgArg 1369
QY 1662 AATGACCTTCAGACACCTTCGACGCGCAGCGGCAATCCGCGCTCGACCTTCGACCGC 1721
Db 1370 GlyArgProAlaArgArgProAlaGlyAlaArgHisAlaGlyValAlaGluPro-----1386
QY 1722 CGAGCTGGGCGCATGCGCCCTGAGCTACGAGAACT 1757
Db 1387 -----GlyProAlaArgArgProGlyArgThrAla 1396

RESULT 7
US-10-366-683-26129
/ Sequence 26129, Application US/10366683
/ GENERAL INFORMATION:
/ APPLICANT: Rubinfeld, Marc J.
/ APPLICANT: Nollring, Jork
/ APPLICANT: Delougnery, Craig
/ APPLICANT: Bush, David
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: PATNO3-04
/ CURRENT APPLICATION NUMBER: US/10/366,683
/ CURRENT FILING DATE: 2003-02-13
/ PRIOR APPLICATION NUMBER: 09/252,991
/ PRIOR FILING DATE: 1999-02-18
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 26129
/ LENGTH: 885
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-10-366-683-26129

Alignment Scores:
Pred. No.: 5,37e-08 Length: 885
Score: 434.00 Matches: 214
Percent Similarity: 34.66% Conservative: 29
Best Local Similarity: 30.53% Mismatches: 248
Query Match: 13.16% Indels: 210
Gaps: 37

US-10-009-782-1 (1-1758) x US-10-366-683-26129 (1-885)
QY 57 CTTGACGCTGCTGCTCGCGGCGG-----CACCTCATTCAGCGCAGCAACACCCGGG 110
Db 37 LeuArgProAlaAspArgGlyArgLeuIleHisArgHisAlaGluPheGlnArgPro---55
QY 111 GGGGGGCGCGCACTGGGCGCTGCGCGGAGCCGATCGCGCCATCGCGCATGTGCGGA 170
Db 56 -----ProGluArgArgHisAlaArgProAlaArgAlaGlyValPro 68
QY 171 GCGCCCGCGCACACCCGCGTGCAGCGCTGCGGCGCTGTGTGTGCGCGCCCGCTTCATCGA 230
Db 69 ArgArgArg-----ProGlyAlaGlnArgTyrProAlaGlyLysArgArgThrAspArg 86
QY 231 CTCGCACACCGACAGACAACTACTGCTCAGAGCGTCCGAGACATACGCCCAAGATCTC 290
Db 87 ThrArgHisProSerArgTyrProAlaProArgArgGlnArgArgValaPheGlnArgSp---105
QY 291 GCAGGGGCTACACCGAGTGTGCACGGGCGCAATTCCGCGATCAACCTGGCGCCGCTGGCGCA 350
Db 106 -----ArgHisArgProGlyHisProAlaGlnGlnHisGln-----GlyIle 119
QY 351 GCGCACCGCGCGCGCGCGCTGAGCACTGCTGCAGACAAAGCGCGCTTACCTTTCGAGCG 410
|||||

```

Db 120 ArgArgProLeuProAlaArgProAlaGlnLysArg-----CysArgAla 135
 QY 411 CTTGCGCCGACTACTGGAGCGGTTGGCGCCACCGCCGCGCCGT-----CAAGCG 461
 Db 136 ValArgArgArgAlaGlyArg-----ArgSerSerAlaGlyArgGlyMetGlyArgGle 154
 QY 462 CGCCTGTATGTGGGGCCATTCAACGCTGCGCGCGGTCAATGCCGAGACTTGACCGCGC 521
 Db 155 AlaLeuAlaGlnLysProArgProGlnGlyArgArgGlnHisArgAsp-----Arg 171
 QY 522 CGGCACCGAGAGAAATCGCGCCAT-----GCGGACCTGGCCGAGAGAAC 569
 Db 172 ProGlnAlaAlaAlaHisArgAlaHisProTyrLeuArgAlaGlyPro----- 187
 QY 570 CATGGCAGCGCGCCATCGGCATTTGACCGCGCCCTTACCGCCCGCCGCGCGC 629
 Db 188 -----ArgArgArgLeuValGlnProAspAlaProArg 198
 QY 630 CACCCACGAGAGATATGAGGTGTGCGCGCGCGGTGAGCGCGCATGGCGCATCTACG 689
 Db 199 His-----AspGlnProGlyArgPro----- 205
 QY 690 CACCCACATGCGGAGAGAGCGGACACATCGTGGCGCGGTGAGAGAACTTCGCGAT 749
 Db 206 -----ArgCysHisArgGlyGlyProAlaGlnProGlyAla 218
 QY 750 CGCGCGCGAGCT-----GAGCGCGCGGTGGTGAATCGCGCAACAGTCAATGGCGCAC 806
 Db 219 LeuProArgSerValGlyArgProAlaGly-----SerProProGlyLeuSerProAla 236
 QY 807 ---CAATTTCGGCGCGTCCGCGAGAGCGTGGCGGTGATCGAGCGCGCAT----- 854
 Db 237 GlyGlnGlyGlnProArgArgGlnProAlaProAlaGlnLysCysProArgArgGlnProPro 256
 QY 855 -----GCGCGCGCGAGAGCGTCTC 872
 Db 257 GlyGlyArgCysProArgProCysArgAlaLeuProGlnProGlyGlnProGlyArgGly 276
 QY 873 GCTGACCGCGATCCCTACGTGGC-----CGG 899
 Db 277 AlaAlaArgArgArgArgArgGlyProValArgArgProPheArgCysThrAlaLeuArg 296
 QY 900 CTCACACATGCTCAAGCAGGAGCGCGTGTGTCGCGCGCGAGCGCACCATCAGCAGCTG--- 956
 Db 297 ArgGlnArgAlaAlaAlaLeuProArgSerValAlaGlnArgAsnProAlaProLeuArg 316
 QY 957 -----GTCAAGCGCTTCCCGCACTAGCGCG-----GCGCGACGTGATGAAGTGC 1004
 Db 317 AlaAlaThrValAlaAlaGlyProAlaArgGlnArgHisArgArgProAlaArgLeuVal 336
 QY 1005 GCGCGAGCGCGGCAATCCAAATCAAGTACGAGCTGTGCGCGAGCTGAGCGCGCGCGCAT 1064
 Db 337 GlyThrAlaArgProValProProAlaArgHisAla--AlaAlaAlaGlyArgThrSer 355
 QY 1065 CTACTCATGATGAGCAGACCGGAGCGAGCGCAT-----CTGCGCGTTCGCGCGCGCAT 1121
 Db 356 ArgArgProValProArgLeuArgArgGlyArgHisArgProGlyAlaAla----- 372
 QY 1122 GATCGCGCTCGAGCGCGCTGCGCA----- 1145
 Db 373 GlyArgLeuArgProProAlaAlaGlySerProGlyProGlyThrAlaValAlaGlyPro 392
 QY 1146 CGAGGAGCGCGCCATCCGCGCGCTGTGGGCGACCTTCCGCGGTGTGCGCGCATATGC 1205
 Db 393 AlaGlnArgProAlaAlaGlyAlaAspAlaHisLeuAsp--GlyLeuArgLeuArgPro 411
 QY 1206 GCGCGACTGTGGCGCT-----GTTCCCGGTGAGAGCGCGGTATGGAAGTGAACCGCGCT 1259
 Db 412 AlaArgProGlyProAspHisValAlaValGlnSerArgProGlnAlaAspProArgPro 431
 QY 1260 GACCGCGCGCGCTTGGCGCTGCGCGCGCGCGCGAGCTGACAGCGCGGTACTCGCGCA 1319
 Db 432 AlaArgArg-----ArgProGlyArgHisAlaGlyAlaAlaArgArgProProAlaAlaArg 449

QY 1320 CTTGTGTGTGTGACCGCGCACGCT----- 1346
 Db 450 ProGlnArgProAlaArgProGlyThrGlyAlaCysGlnProAlaMetAlaAlaGlyThrArg 469
 QY 1347 -----GCCGATACCGC----- 1358
 Db 470 ProValValAlaAlaProAlaGlnArgArgArgThrGlnProGlyLeuArgProAspPro 489
 QY 1359 ----- 1373
 Db 490 AlaGlyTyrGlyProAspArgProAlaSerArgThrAlaGlyTyrLeuAlaGlyPro 509
 QY 1374 TACCGAGCGCGCGCGCGCATTCCTGATCGATCAACGCGCGCGGTGCGCAGA 1433
 Db 510 ValAlaAlaGlyArgProGlnProPheArgProGlnHisProAlaAspGlyLysGlyArg 529
 QY 1434 GCAGCGCTTCACCGCGCGCATGCGCGCGCGGTGTCGCAAG-----CACGCGCGC 1484
 Db 530 LeuArgThrSerArgAlaLysAlaLeuProAlaArgArgAlaValProHisGlySer 549
 QY 1485 CTGAGCGCG-----GCCAGCGCTTACATCCGCGGTGAGAG 1523
 Db 550 ArgAlaArgProGlySerGlyGlnArgGlnAlaGlyProGlyAlaArgArgArgAlaGln 569
 QY 1524 GCGCGCGTGCAGCGCGCT-----CCCAACCGTGA-----CGCAA 1559
 Db 570 ArgAlaCysProAlaAlaAlaArgProArgProAspLeuValArgAlaThrAlaLeuArgGln 589
 QY 1560 CCGCTACATGCGCGCTCCCTC-----CGCTCGCATACGCGCGCGCGCATATGCT 1610
 Db 590 ProAlaHisArgArgThrLeuValGlnArgAlaAlaAlaGlyGlyProAlaGlyArgVal 609
 QY 1611 GCGCAAGAGAGTGAAGCT 1670
 Db 610 GlyGlyArgGlnAlaGlyArgGlnProAlaArg-----ProGlyArgArgPro 625
 QY 1671 GCA-----AGACGTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1721
 Db 626 AlaHisLeuGlnArgArgArgAlaGlnLeuSerArgGlyProGlyProGlnProAlaGlnGly 645
 QY 1722 CGAGCTGG-----CCAGATCGC 1739
 Db 646 GlnValGlyGlnPheProAlaGlnLeuLeuHisGlyLeuProAlaArgHisProAspArg 665
 Db 666 Pro 666
 QY 1740 CCT 1742
 Db 666 Pro 666
 RESULT 8
 US-10-419-128-26129
 ; Sequence 26129, Application US/10419128
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/10/419,128
 ; PRIORITY FILING DATE: 2003-04-21
 ; PRIOR APPLICATION NUMBER: US/09/252,991
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 26129
 ; LENGTH: 885
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-10-419-128-26129
 Alignment Scores:

Pred. No.: 5.37e-08
 Score: 434.00
 Percent Similarity: 34.66%
 Best Local Similarity: 30.53%
 Query Match: 13.16%
 DB: 6
 Length: 885
 Matches: 214
 Conserved: 29
 Mismatches: 248
 Indels: 210
 Gaps: 37

US-10-009-782-1 (1-1758) x US-10-419-128-26129 (1-885)

```

QY 57 CTTGACCTGCTGCTGGGGGGG-----CACCTCATGACGCGACACACCGGG 110
DB 37 LeuArgProAlaAspArgGlyLeuGlnHisArgHisArgLeuPheGlnArgPro--- 55
QY 111 GCGGCGCGCGACCTGGCGCGTGGCGGCGGCGACCTGCGCCATCGGCATCTCGGA 170
DB 56 -----ProGlnArgHisAlaArgProAlaArgValArgValPro 68
QY 171 GCGGCGCGCGACACCGGGGTCGACGCTGGCGCGCTGGTGGCGCGCGCTTCATCGA 230
DB 69 ArgArgArg-----ProGlnAlaGlnArgTyrProAlaGlyLysArgTyrHisArg 86
QY 231 CTGCGACACCGCGACGACACAACTACCTGCTCAGCGCTCGGACATGACCGCAAGATCTC 290
DB 87 ThrArgHisProSerGlyProAlaProAlaArgGlnArgValArgValArgValArg 105
QY 291 GCAGGGGTCACACGCGTGTCTACGGGCAATTCGCGCATCAGCGCTGGCGCGCTGGCGCA 350
DB 106 -----ArgHisArgProAlaHisArgProAlaGlnGlnHisGln-----GlyGly 119
QY 351 CGGCAACCGCGCGCGCGCGCGCTGGACCTGCTGACGACGAGCGCGCTTACCGTTTCAGCG 410
DB 120 ArgArgProLeuProProAlaArgProAlaArgProAlaGlnArg-----CysArgAla 135
QY 411 CTTCGCGGACTACCTGACGCGCTTGGCGCGCGCGCGCGCGCT-----CACGC 461
DB 136 ValArgArgArgAlaGlyArg-----ArgSerSerAlaGlyGlyArgGlyMetGlyArg 154
QY 462 GCGCTGTATGGTGGCGCTTCACGCTGCGCGCGCGCGCTTCATCGCGACTTCGACGCG 521
DB 155 AlaLeuAlaGlyGlnProArgProGlnGlyArgGlnHisArgGlnHisArgGlnHis 171
QY 522 CGGCGCGCGACGAGAAATCGCGCGCAT-----GCGGCGACTGGCGCGCGAGAGAGC 569
DB 172 ProGlnAlaAlaAlaHisArgAlaHisArgProTyrLeuArgAlaGlyPro----- 187
QY 570 CATGGCGCGCGCGCGCATTCGACCTGCGCGCGCTTACCGCGCGCGCGCGCGCG 629
DB 188 -----ArgArgArgLeuValGlnProAlaArgProArg 198
QY 630 CACCGCGAGAGATCATCGAGGTCGCGCGCGCTGACGCGCGCATGCGCGCATCTACGC 689
DB 199 His-----AspGlnProGlyArgPro----- 205
QY 690 CACCGCGATCGCGCGAGAGAGACATCGTGGCGCGCGCTGAGAGAACTTCGCGAT 749
DB 206 -----ArgCysHisArgGlyGlyProAlaGlyGlnProGlyAla 218
QY 750 CGGCGCGGAGCT-----GGAGCTGCGCGCTGATCTCGACACCAAGTATGGCGCGCG 806
DB 219 LeuProArgSerValGlyArgProAlaGly-----SerProGlnGlyLeuSerProAla 236
QY 807 -----CAATTCGCGCGCTGCGCGAGAGCGCTGCGCTGATCGAGCGCGCGCAT----- 854
DB 237 GlyGlnGlyGlnProArgArgGlnProAlaProAlaGlyCysProArgArgGlnProPro 256
QY 855 -----GGCGCGCGAGAGCGCTC 872
DB 257 GlyGlyArgCysProArgProCysArgAlaLeuProGlnProGlnProGlnProGlyArgGly 276
QY 873 GCTGAGCGCGTATCCCTAGTGG-----CGG 899
DB 277 AlaAlaAlaArgArgArgArgGlyProValArgArgProPheArgCysThrAlaLeuArg 296

```

```

QY 900 CTCACCATGCTACAGCAGAGACCGCGCTGCTGGCGGACGCACATCATCCTG--- 956
DB 297 ArgGlnArgAlaAlaAlaLeuProArgSerValAlaGlnArgGlnArgProAlaProLeuArg 316
QY 957 -----GTGACGCGCTTCCCGCAACTGACGCG-----GCGGACCTGATGAATCCG 1004
DB 317 AlaAlaThrValAlaAlaGlyProAlaArgGlnArgHisArgArgProAlaArgVal 336
QY 1005 GCGCGAGCGCGCAATTCACAGTACGACGCTGGCGCGCGCGCGCGCGCGCGCAT 1064
DB 337 GlyThrAlaArgProValProProAlaArgHisAla-----AlaAlaAlaGlyArgHis 355
QY 1065 CTACTCATGATGAGCAACCGGACGCTGACGCGCAT-----CTGCGCTTGGCGCGCGCAT 1121
DB 356 ArgArgProValProArgLeuArgGlyArgHisArgProGlyAla----- 372
QY 1122 GATCGCTCGACGCGCTGCGCG----- 1145
DB 373 GlyArgLeuArgProProAlaAlaGlySerProGlyTyrAlaValAlaGlyPro 392
QY 1146 CGACGAGCGCGCGCATTCGCGCGCTGGCGCGCTTCCCGCGCTGGCGCGCATATGC 1205
DB 393 AlaGlnArgProAlaAlaGlyAlaAlaSerAlaHisLeuAsp-----GlyLeuArgLeuArgPro 411
QY 1206 GCGGCGCGCGCGCT-----GTTCCGCTGAGAGCGCGGTATGAGATGACGCGCT 1259
DB 412 AlaArgProGlyProAspHisValAlaValGlnSerArgProGlnAlaAspProArgPro 431
QY 1260 GACCGCGCGCGCTTGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1319
DB 432 AlaArgArg-----ArgProGlyArgHisAlaGlyAlaArgArgProProAlaAlaArg 449
QY 1320 CCGTGTGTGTTCACCGCGCGCGCGCT----- 1346
DB 450 ProGlnArgProAlaProGlyThrGlyAlaCysGlnProAlaMetAlaAlaGlyThrArg 469
QY 1347 -----GGCGCATACCGC----- 1358
DB 470 ProValValAlaAlaProAlaGlyArgArgThrGlnProGlyLeuArgProAspPro 489
QY 1359 -----CACCTTGCAACCGC 1373
DB 490 AlaGlyTyrGlyProAspArgProAlaSerArgArgThrAlaGlyTyrLeuAlaGlyPro 509
QY 1374 TACCGAGCGCGCGCGCGCATTCGCTGATGATGACGCGCGCGCGCGCGCGCGCG 1433
DB 510 ValAlaAlaGlyArgProGlnProPheArgProGlnHisProAlaAspGlyLysGlyArg 529
QY 1434 GCAGCGCTTACCGCGCGCATGCGCGCGCGCGCTGCGCGCG-----CACGCGCGC 1484
DB 530 LeuArgThrSerArgAlaAlaAlaLeuProArgArgAlaAlaValProHisGlySer 549
QY 1485 CTGACCGCG-----GCGCGAGCGCTTACCAATCCGCGCGCGCGCGCGCG 1533
DB 550 ArgAlaProGlySerGlyGlyArgGlnAlaGlyProGlyAlaArgArgArgAlaGln 559
QY 1524 GCGCGCGCGCGCGCGCGCG-----CCGACCTGGA-----CGGAA 1559
DB 570 ArgAlaCysProAlaAlaAlaArgProArgProAspLeuValArgAlaThrAlaLeuArgGln 589
QY 1560 CCGCTACATGCGCGCGCTGCT-----CGCTCGCAATAGCGCGCGCGCGCGCGCG 1610
DB 590 ProAlaHisArgArgThrLeuValGlyAlaArgAlaAlaAlaGlyLysProAlaGlyArgVal 609
QY 1611 GCGGAGAGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1670
DB 610 GlyGlyArgGlyAlaGlyGlnProAlaTyr-----ProGlyArgArgPro 625
QY 1671 GCA-----AGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625
DB 626 AlaHisLeuGlnArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 645
QY 1722 CGAGCTGGG-----CGAATCGC 1739

```

```

Db      646  GLUVALGlyGlnPheProAlaGluLeuHisGlyLeuProAlaArgHisProAspArg 665
QY      1740 CCT 1742
Db      666  Pro 666

RESULT 9
US-10-366-683-19446
Sequence 19446, Application US/10366683
GENERAL INFORMATION:
APPLICANT: Rubenfeld, Marc J.
APPLICANT: Neilling, Jork
APPLICANT: Delouherie, Craig
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: PAT003-04
CURRENT APPLICATION NUMBER: US/10/366, 683
PRIOR FILING DATE: 2003-02-13
PRIOR APPLICATION NUMBER: 09/252,991
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 19446
LENGTH: 937
TYPE: PR
ORGANISM: Pseudomonas aeruginosa
US-10-366-683-19446

Alignment Scores:
Pred. No.: 5,75e-08 Length: 937
Score: 433.00 Matches: 234
Percent Similarity: 34.29% Conservative: 41
Best Local Similarity: 29.18% Mismatches: 257
Query Match: 13.13% Indels: 270
Gaps: 42

US-10-009-782-1 (1-1758) x US-10-366-683-19446 (1-937)
QY      6  CCACTGATGCGGAGAGAGATTCATGCGATTCGATCCAGCCCTTCAGCT 65
Db      112  ProLeuProGlnArgThrArgHisProAlaHisArgPro-----GlyThrGlyArgPro 128
QY      66  GCTGCTCGCGGGGCGGACCTCAT-----CGACGGGCAACAACCCGGG 110
Db      129  ---AlaHisProGlnHisProHisLeuPheGlyProGlyArgHisArgHisProGly 147
QY      111  GCGGCGCGGACCTGGGGGTGGCGGCGGACCGCATCGCGCGCATGCTGCGGA 170
Db      148  Ala-----GlyArgGlyValArgProAlaArgGln----- 157
QY      171  CGCGCGCGGCGACACCGGGGTGGA-----CGTGTGGGGCTGGTGGTGGCGCGGCTT 224
Db      158  -----ProGlnHisLeuAlaArgProGlyProGlyGlnGlnArgGlyArg 172
QY      225  CATGCACTGCGACACCGGACGACCACTACTGCTGCGGCGGTGGCGACGACGCGCA 284
Db      173  AspArgProArgHisProHisArgGlnArgValAlaGluArgGlyAlaSerAspSerArg 192
QY      285  GAT----- 287
Db      193  GlnArgGlyAlaValProProArgGlyAspGlyThrValAspArgLeuProArgPro 212
QY      288  -----CTCGAGGCGCGTACACCGGTGGTGC-----GGGGAATTGGGGCAT 329
Db      213  GlyProArgGlyGlyGlnGlySerGlyAspHisArgArgThrValAlaArgLeuProArg 232
QY      330  CAGCTGGCGCGCGTGGCGGCGACGACCGCGCGCGCGCGCTGGAAGCTGCT----- 380
Db      233  IleProGlyThrGlyAlaThrHisArgArgProAspArgArgProArgProAlaLeuGly 252
QY      381  -----GGAGAGGCGGCTTACCGTTGAGAGGCTTGGCGGCGCTGACTGCTGA 428

```

```

Db      253  GlyThrAlaGlyGlyArgArgGlyGlyLeuArgAlaArgThrArgGlnGly 272
QY      429  CGGTTGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 485
Db      273  ArgLeuProGlnGluAlaAlaAlaAlaAlaArgArgGlyArgLeu----- 286
QY      486  GGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 542
Db      287  AlaGlnArgAlaHisAla-----ArgGlnArgGlyAspThrArg 301
QY      543  -----GGCCAT-----GCGGACCTGGCGGAGAGGCGGCGGCGGCGGCGGCGG 581
Db      302  GlyProGlnHisLeuProAlaAlaProAspGlnArgAlaGlnArgArgGlyGlnLeu 321
QY      582  ---GGCCATGCGGCAATTC-----GACGCGCGCTTCTACCGCGCGCG 620
Db      322  LeuArgHisArgSerLeuArgProAlaLeuGlnGlyGlnArgArgGlyGlyArg 341
QY      620  ----- 620
Db      342  LeuLeuGlyArgLeuGlnArgArgProGlnGlyGlnValGlnLeuHisArgAlaGlyGly 361
QY      621  -----CGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 656
Db      362  AlaAspSerGlnValAlaAlaArgProGlnHisArgGlyGlyThrArgGlyArgLeu 381
QY      657  ---CGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 671
Db      382  HisProAlaAlaAspArgGlnPheLeuAspAlaProAlaArgGlnAspLeuProArgArg 401
QY      672  -----CGATGG 677
Db      402  GlyLeuValArgLeuArgLeuGlyAlaGlyValAspArgLeuArgValGlnProAlaVal 421
QY      678  CGGCACTGTACGCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 737
Db      422  GlnHisLeuValGlnProAsp---ArgArgArgValAlaGlyArgArgArgAlaGlyGly 440
QY      738  AACCTTCCGATCGG-----CGGAGCGTGGAGCGGCGGCGGCGGCGGCGGCGG 779
Db      441  Gly-----HisArgAlaValHisArgGlyProArgAlaGlyArgGlyLeuAspAla 458
QY      780  -----CTGCGACACCAAGCTCAT 797
Db      459  GlnAlaAlaProAlaIleProAlaAspHisArgArgAlaGlyLeuSerAlaGlnGlyVal 478
QY      798  GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 857
Db      479  AspProArgAlaLeuLeuGlnArgAlaAlaGlyThrAlaGlnAlaAspProArgArgPro 498
QY      858  GCGCGAGGACGTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 917
Db      499  CysProProArgLeuProGlyLeuArgSerProGlyAspArgGlnGlnHisProArgPro 518
QY      918  GACCGCGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 977
Db      519  GlyArgLeuAlaAlaGlyArgGlyAlaLeuArgAlaProGlyArgAlaLeuProLeu 538
QY      978  GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1028
Db      539  ProArgArgProAlaGlyArgLeuGlnGlyArgArgAlaGlnGlyLeuArgProGlyProArg 558
QY      1029  -----GAGCTGTGTCGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1079
Db      559  GlyGlyGlyArgArgGlyGlyArg-----GlyAspArgArgArgLeuLeuArg----- 574
QY      1080  GGAACCGGACGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1139
Db      575  ArgProArgLeuAlaGlnAlaHisGlyAla-----AlaLeuArgArgPro 589
QY      1140  GCGCGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1193
Db      590  AlaAspArgArgGlyGlyAlaValAlaAla-----GlyLeuProArgProAlaGln 606

```

[illegible]

```

QY 621 -----CGCCGCGCCACCACCGAAGAGATCATGAGTGTG----- 656
DB 362 Alaaspserglnvalalaargprogluhsargleuaglythrlarglythargleu 381
QY 657 -----CGGCGCGGTGAGCGC----- 671
DB 382 HlsprolaalaasparaglnphleuaspalaProlaaargluasprleuProaarg 401
QY 672 -----GCATGG 677
DB 402 GilyleuvalargleuargleuaglyalaaglyalaasparargleuarglyalnProalaVal 421
QY 678 CGGATCTAGTACCCACCACTGCGCGAAGCGAGACACATGCGCGCGTGGAGGA 737
DB 422 GlnhlsleuvalalnProaP---Argargargvalalaaglyargargarglaaglygly 440
QY 738 AACCTTCGCGTGG-----CGCGAGCTGAGAGTGGCGGTGTGAT--- 779
DB 441 Gily-----HlsarglaValaHlsargglyProaarglaaglyarglyleuaspala 458
QY 780 -----CTCGACCAACAAGTCAT 797
DB 459 GlnalaalaProalaileProalaasphlsargarglaaglyleuaserlaaglnaglyVal 478
QY 798 GGGCGAGCCCAATTTGGCGCGCTCGCGAAGACGCTCGCTGATGAGGCGCATGGC 857
DB 479 Aspproaarglaaleuaglnaarglaalaaglythrlaaglylnlaasproaargargpro 498
QY 858 GCGCCAGGAGCTGCTGCTGAGCGCGGTACCTACGTGGCGCGCTCCACATGCTCAAGA 917
DB 499 CysProaargleuProaileuargserProaileuarglnlaasproaargpro 518
QY 918 GAGCGCGCTGCTGCTGCGCGAGCACCATCATCTGTCAGACGCTTCGCGCAACT 977
DB 519 GilyargleuvalalaaglyargglyalaaleuargalaProaileuargleu 538
QY 978 GAGCGGCGCGACTGATGAAGTGGCGCGGCGGCGCAATCCAGTA----- 1028
DB 539 ProaargargproaileaglyargleuaglnlaargargalaaglyleuargProaileuarg 558
QY 1029 -----CGAGCTGTGTGCGCGGCTGAGCGCGCGGCGCATCTTATCATATGA 1079
DB 559 Gilyaglyargargglyarglyarg-----Glyasparargargleuarg----- 574
QY 1080 CGAACCCGAGCTGAGCGGCGCATCTGCGCGCTGCGCGCGCATGATGCTGCGAGCGCT 1139
DB 575 ArgproaargleuvalaaglnalaHlslyala-----Alalaargargpro 589
QY 1140 GCGCGCAGCAGCGCGCGCATCTGCGCGCTGCGCGCATCTCCGCGGTGCT----- 1193
DB 590 AlaasparargarglyalaValaala-----GlyleuProaargProaileuarg 606
QY 1194 -----GGGCACTATGCGCGCGCA---CCTGGGCTGTCTTCCGCTGAGAGCGGGCT 1241
DB 607 ArgleuaglnalaaleuargargValaGlnlyleuProaileuargHlslyHls 626
QY 1242 ATGGAATGATGAGCGCGCTGACCGCGC-----GCGCTTCGG 1277
DB 627 ProaileuargproaileuargargphlsarglaargHlsHlsasphlsasparProaileuarg 646
QY 1278 CCGGCGCGCGCGCGCGAGCTGCGCGGTCTTCTGCGCGA-----CCT 1322
DB 647 Alaglyarglaaglu---MetalaaglyMetvalaHlsHlsargglyargarglaaglyPro 665
QY 1323 GGTGTGTGTGACCGCGCGCGCGCGATG---CGCACCTTCGACACACCTTAC 1376
DB 666 AlaglyvalaarglaaglyProaileuargleuarglaaglnleuargglnaglygly 685
QY 1377 CGAGCGCGCGCGCGCATCTCTGTCGTCACAGCGCGCGCGGTGCGAGGA--- 1433
DB 686 AsparaileuargHlsargleuaglnlaalaaleuProaileuargleuargargarg 705
QY 1434 -----GACGCGCTTACCGCGCGCATGCGCGCGCGGTCTCGCACGCGACGGC 1481

```

```

DB 706 ProaasphlsaglnalaalaargargargproailealaargProaileuargarglnPro 725
QY 1482 CGGCTAGCGCGCGCGCGCATCTTACATCGCGGTGAACGGGCGCGCTGCCGCCCT 1541
DB 726 Pro---AspproaargproaileuargleuargglyargleuvalalaVala----- 741
QY 1542 CCCAACCTTGACGCAACCGCTACAT----- 1568
DB 742 -----AsparargargargproaileuargleuHlsProaileuarglaalaaleuValgly 759
QY 1569 GCGCGCTCTCTCGCGCGA----- 1589
DB 760 GilyasparHlsarglaaglnalaargargProailealaalaasprleuargargProala 779
QY 1590 TACGCGCGCGCGCATCTGTCGCGCAAGATGATGGCGCGCGCTGCGCGCGAGG 1649
DB 780 AlaglyProaileuargleuaglnlaasparProaileuargProaileuargHlsarg 799
QY 1650 CAAGCGCGCGA-----AATGACCTGCAAGACGTGCGCA----- 1685
DB 800 ArgargProaileuargleuValleuargargProaileuargProaileuargHls 819
QY 1686 -----GCGCAGCGCATCGCGGTCTGAC----- 1709
DB 820 CysglnaglyalaValaargleuargHlsProaileuargHlsProaileuargHls 839
QY 1710 -----CCTGTCACAGCGCGAGCTGCGCGCATCGCGCTGAGCTCAAGA 1751
DB 840 AsparaileuaglnProaileuargProaileuarglyglyProaileuargglyserleuarg 859
QY 1752 GAGCT 1757
DB 860 Aspara 861

RESULT 11
US-10-366-683-30843
Sequence 30843, Application US/10366683
GENERAL INFORMATION:
APPLICANT: Rudenfeld, Marc J.
APPLICANT: Nolling, Jork
APPLICANT: Deloughery, Craly
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: PAT03-04
CURRENT APPLICATION NUMBER: US/10/366,683
CURRENT FILING DATE: 2003-02-13
PRIOR APPLICATION NUMBER: 09/252,991
PRIOR FILING DATE: 1999-02-18
NUMBER OF SEQ. ID NOS: 3142
SEQ. ID NO 30843
LENGTH: 663
TYPE: PRN
ORGANISM: Pseudomonas aeruginosa
US-10-366-683-30843

Alignment Scores:
Pred. No.: 6.37e-08
Score: 432.50
Percent Similarity: 33.51%
Best Local Similarity: 27.52%
Query Match: 13.11%
DB: 6
Gaps: 40

US-10-009-782-1 (1-1758) x US-10-366-683-30843 (1-663)
QY 32 CGATGCCCCAATCCGATTCACGAC-----CCTGACCTGCTGCGCGCGCGACCC 85
DB 4 ProaileuargproaileuargargphlsarglaargHlsHlsasphlsasparProaileuarg 23
QY 86 TCATGACGCGCGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 145

```

```

Db 24 TrpSerThrThrThrPro-----ProAlaAspArgSerCylValAlaThrAla 39
QY 146 TGGCCGCGCATCGCGATCTGTGAGACGGCGCGACACCGCGGTGAGTGTGGGCC 205
Db 40 ArgValProAla-----CysArg-----ArgValProIleArgCysAlaAla 54
QY 206 TGGTGTGGCGCGCGCTTCATGACTGCGACACCCACAGACAGACAACTACTGCTCAGGC 265
Db 55 TrpProGlyArgProMetSerSerThr----- 63
QY 266 GTGCGACATGACGCCCAAGATCTGCGAGGGCGGTACACGAGTGTGTCAGGCAATTGCG 325
Db 64 -----SerArgArg-----ProGlyTrpSerProAlaValPro 74
QY 326 GCATCAGCC-----TGGCGC----- 346
Db 75 AlaGlyAlaCysCysTrpIleThrArgAsnValAlaIleThrArgSerAlaThrAspAlaTrp 94
QY 347 CGCAGCGCCAAACCGCGCGCGCGCGCTGACCTGTCAGAGGAGGCGGTCTTACCGTTGCG 406
Db 95 SerSerAlaProAlaProAlaAsnCysCysLeuPheProAlaGlyProSerProAla 114
QY 407 AGCGCTGCGCGCTACTGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 463
Db 115 SerProAlaProAlaAlaCysArgArg-----TrpProArgAlaAlaCysHisTrpProAla 133
QY 464 CCGTATGCTGGCGCATCTACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
Db 134 SerAlaTrpTrpTrpLeu-----AlaProAlaArgCysCysTrpProAlaProPro 150
QY 524 CCACGCGACGAGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571
Db 151 AlaProAlaSerAlaGlyArgAlaCysCysAlaSerProAsnArgArgArgLeuPro 170
QY 572 TGGCCGCGCGCGCGCATCGCGATTCGACGCGCGCGCTTCTACG-----CGCGCGCGCGCG 625
Db 171 TrpProSerProTrpAlaSerArgAlaGlyProAlaSerCysGlyArgProProAla 190
QY 626 ----- 658
Db 191 CysSerProValAlaThrAlaProThrAlaThrCysSerProProSerAlaArgSerAla 210
QY 650 -----AGGTGTGCC----- 658
Db 211 TrpLysProCysAlaCysAlaLeuValAlaGlySerAlaArgSerProAlaAsnAlaTrp 230
QY 658 ----- 658
Db 231 ProAlaAlaSerAlaTrpCysProThrCysAsnSerAlaArgProSerAlaIleAlaSer 250
QY 659 -----GGCGCGTGAAGCGCGCGCATGGCGCGCATCTACGCCA 691
Db 251 ThrAlaArgArgSerProSerThrAsnGlyArgProAlaAlaCysProThrThrThrPro 270
QY 692 CCCACATAGCGCG-----ACGAAGCGGACACATCGTGGCGCGCGCGGAGAAACCTTCCGCA 748
Db 271 ProIleAsnAlaProAlaSerAlaAlaAlaAsnTrpProTrpTrpLysAlaProSerPro 290
QY 749 TCGCGCGCGAGCTGAGACGTGGCGGTGTGATCTCGCACCAAGATCATGGCGCGCGCA 808
Db 291 AlaMetProProSerThr----- 296
QY 809 ATTTGGCGCGCTGCGCGCGAGACGCTGC-----CGCTGATCAGAG 847
Db 297 -----SerAlaThrArgLeuAlaAlaCysGlyHisAspGlyAspAlaGlyArgAlaSerGln 315
QY 848 CCGCCATGGCGC----- 859
Db 316 ThrProTrpArgAspThrSerProCysValProAsnCysValAsnTrpArgArgArgIle 335
QY 860 -----GCCAGAGCGTCTGCTGCGAGCGCTATCCTTACGTGGCGCGCGCTCA 904
Db 336 ProTrpSerAlaValAlaAlaLysThrSerProTrp-----ProArgTrpProGlyMet 352

```

```

QY 905 CCATGCTCAACAGACCGCGCTGC-----TGTGGCGCGGACGACCACTCAATCA 952
Db 353 ProAlaGlyProArgProSerCysThrAlaValAlaAlaTrp-----AlaProAlaArg 370
QY 953 CCTGTGCAGACCGCTTCCCGAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1012
Db 371 GlyIleThrAlaAlaAlaAlaProArg-----AsnSerCys-----SerAlaGlyArg 386
QY 1013 GCGGCAATCAAGTACAGAGTGTGGCCCAAGCTGCA-----GCCGCGCG 1057
Db 386 ValArgArg-----ArgValArgArgSerAlaArgAlaTrpArgProTrpProAlaGlyArg 405
QY 1058 GCGCCATCTACTCATGATGAGACGACCGACGCGCGCGCGCGCGCGCGCGCGCGCG 1111
Db 405 GalThrProAlaSerAlaAlaAlaProArg-----CATCTGCGCTTGC 1111
QY 1112 GCCGCAATGATCGGCTCGACGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1171
Db 425 AlaAspHisProAlaAlaAlaTrpValSerAlaAlaAlaArgArgThr-----SerSerAlaProIle 444
QY 1172 G-----GGGACCTTCCCGCGCGGTGCTGGGCGCACTATGCGCGCGCGCGCG 1214
Db 444 AlaGlySerAlaProGlySerCylThrAlaProArgCysHisProValArgLysAspGly 464
QY 1215 -----GGGCGCTG-----TCCCGCTGAGAGCGCGGTATGAGATGACCGCGCG 1258
Db 464 ValAlaGlyProAlaSerThrGlyArgSerArgArgArgTrpAlaAlaAlaProAlaAlaArg 484
QY 1259 -----TACCAGCGCG----- 1267
Db 484 GalAlaArgThrGlyLeuArgArgThrAlaAlaAlaAlaCysArgTrpProProAlaProGly 504
QY 1268 -----CGGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1293
Db 504 ValArgThrProAlaAlaAlaArgArgArgSerAlaAlaArgThrAlaProArgProAlaIleArg 524
QY 1294 CAGCTGCGAGCGCGGCTACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1353
Db 524 AlaHisArgProSerAlaSerAlaSerAlaThrGlyTrp-----ProGlyProAlaLeuAlaCys 542
QY 1354 ACCGCGACCTTGCAGACACCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413
Db 542 sProAlaAlaAlaGlyArgArgProSerProAlaProAlaProAlaProAlaAla 562
QY 1414 GCGCGCGCGGTGCGCAAGACAG-----CGTTCACCGCGCGCGCGCGCGCGCGCG 1467
Db 562 ArgProAlaAlaProAlaProArgArgArgProGlyArgSerProAlaThrAlaProAlaAlaAla 582
QY 1468 CTCGACGACGACGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1527
Db 582 AlaAlaProAlaGlyProThrProAlaAlaArgProPro-----AlaAlaAlaThrGlyArg 600
QY 1528 GCGTGCAGCGCGCTTCCCAACCTGAGCGCAACCGCTCATGCGCGCGCGCGCGCGCG 1587
Db 600 gProSerAlaProAlaProAlaArgArgArgArgArgArgArgArgArgArgArgArg 620
QY 1588 AATA-----CGGCGCGCAACCGCATTCGTGGCGCAAGAGTATGGCGCGCGCG 1638
Db 620 gCysSerProArgArgProThrProArgSer----- 630
QY 1639 GCGCGCGAGCGCAAGCGCGCGCAAGTACCTGCAAGACGCTGTGCGAGCGCGCGCGCATC 1698
Db 631 AlaProGlyAlaAlaThrGlyAla-----ProThrThrGlyAlaAlaAlaArgPro----- 646
QY 1699 GCGGTCTGACCGCTGTCCAAAGCGAGCTGGCGCGAGATCG 1738
Db 647 ArgArgArgArgAlaAlaAlaAlaProGlyLysArgArgSer 659

```

RESULT 12
 US-10-419-128-30843
 ; Sequence 30843, Application US/10419128
 ; GENERAL INFORMATION:

US-10-009-782-1 (1-1758) X US-10-419-128-30843 (1-663)

[illegible]

```

Db 524 oalnhsatgprroseralaseralathrlgltpr-----ProglyProalaLeuAlaCy 542
QY 1354 ACCGSCACCTTCGAAACCTACGAGCGCGCGCGGATTCATTCGCTGACGTCAAC 1413
Db 542 sPrroAlaAlaAlaGlyAArgArProSerProAlaProAlaProAlaProAlaProAla 562
QY 1414 GCGGCGCGCGCTGTCGCAAGCAG-----CGTTCACGCGCGCATGCGCGCGCGCTG 1467
Db 562 aArgProAlaAlaProAlaArgArProGlyAArgSerProAlaAlaProAlaAlaProAla 582
QY 1468 CTGCGACGACGCGCGCGCTGACCGCGCGCGCGCGCGCTTACATCCGCGTGAACGCGCG 1527
Db 582 eAlaAlaAlaProAlaGlyProAlaProAlaProAlaProAlaProAlaProAlaPro 600
QY 1528 GCGTGGCGCGCGCTTCACACCTTCGACGCAACCGCTACATGCGCGCGCTTCGCGCTGCG 1587
Db 600 gProSerAlaAlaProAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 620
QY 1588 AATA-----CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1638
Db 620 gCySerProAlaArgArProAlaProAlaProAlaProAlaProAlaProAlaProAla 630
QY 1639 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1698
Db 631 -AlaProGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 646
QY 1699 GCGGCTCGACCGCTTCGCAAGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1738
Db 647 -ArgArArgArAlaAlaAlaAlaAlaAlaProGlyLysArArgArSer 659

```

RESULT 13

```

US-10-366-683-21920
Sequence 21920, Application US/10366683
GENERAL INFORMATION:
APPLICANT: Rubinfeld, Marc J.
APPLICANT: Nollind, Jork
APPLICANT: Delouhery, Craig
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: PATH03-04
CURRENT APPLICATION NUMBER: US/10/366,683
PRIOR FILING DATE: 2003-02-13
PRIOR APPLICATION NUMBER: 09/252,991
PRIOR FILING DATE: 1999-02-18
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21920
LENGTH: 822
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-10-366-683-21920

```

Alignment Scores:

```

Pred. No.: 6,64e-08 Length: 822
Score: 431.50 Matches: 223
Percent Similarity: 35.848 Conservative: 34
Best Local Similarity: 31.108 Mismatches: 233
Query Match: 13.088 Indels: 227
DB: 6 Gaps: 42

```

US-10-009-782-1 (1-1758) x US-10-366-683-21920 (1-822)

```

QY 45 CGATTCGACCGCTTCGACTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104
Db 2 ArgArArgProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 18
QY 105 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152
Db 19 ProGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 38
QY 153 CATTGGGATCTGTCTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 212

```

```

Db 39 GlnArArgGlnProAlaArgArGlnAlaGlnProGlyAlaAlaAlaGlnAlaAlaAla 58
QY 213 CGCGCGCGCGCTTCATGACCTTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 245
Db 59 GlnProAlaGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 245
QY 246 ---CGACAACTACTGCTGAG---GCGTGGCGACATGACCGCGCGCGCGCGCGCGCGCG 299
Db 79 GlnArArgArGlnProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 96
QY 300 CACCAAGGCGGTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 344
Db 97 ArgProAlaGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 113
QY 345 GCG-----GACGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 371
Db 114 GlyProGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 133
QY 372 GACCTGCTGACGA-----AGCGCGCTCTACCGCTT 404
Db 134 GlyAlaProGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 153
QY 405 CGAGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 464
Db 154 ArgAlaAlaGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 172
QY 465 CTGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 524
Db 173 -----ArgArArgArAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 183
QY 525 CACCGACGAGGAATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 581
Db 184 SerArgArAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 203
QY 582 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641
Db 204 ProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 213
QY 642 GATCATGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701
Db 214 -----GlyTPrArgProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 223
QY 702 CGACGAGGCGA-----GCACATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 734
Db 224 GlnProAlaArgProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 243
QY 735 GGAACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 791
Db 244 ArgArArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 263
QY 792 GGTCA-----GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
Db 264 GlyLysArGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 275
QY 840 GATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 878
Db 276 AspArgGlnGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 295
QY 879 CGCGTATCCCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 929
Db 296 ArgTPrArgProGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 315
QY 930 GGT-----GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 968
Db 316 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 330
QY 969 CCGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1007
Db 331 ---ArgGlnTPrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 349
QY 1008 -----CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040
Db 350 AspAlaGlnTPrProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 369

```

```

QY 1041 CGAGCTGCA-----GCCGGCGCGGCATCTACTCATGATGACGAACCGACGTGCA 1094
DB 370 SerATGThProGluProGluProGluHisArgHisLeuProProAspLysArgLysAlaAla 389
QY 1095 GCGCATCTGCGGTGCGCGCCGACCATGATCGGCTC-----CGACGGCTCGCGCA---- 1145
DB 390 GtGtAlaValLeuHisTrpProArgArgArgLeuGluGlnHisArgTrpProAlaAlaGln 409
QY 1146 CGACGAGCGCGCGCATCCCGCTGTGGGGACACCTCCGCGGGTGGTGGGACATATGC 1205
DB 410 ArgArgGlnLeuAlaProArgProGluArgArgLeuPro-----Trp 423
QY 1206 GCGGACCTGGGCTGTTCCTGCTGGAGACGGCGGTATGGAAGATGACCGGCTGACGC 1265
DB 424 GlnArgProArgHisGlnPro-----ArgArgLeuArgProGluArg 437
QY 1266 GCGCGGCTTCGGCT-----GCGCGG 1286
DB 438 ArgGlnProArgProArgGlnArgArgGlnProArgArgArgGlnCysHisAlaArg 457
QY 1287 GCGGCGGAGCTGCAGCGCGGTA-----CTTCGCGCACTGTGTGTGT 1331
DB 458 ArgArgGlnProAlaGlnArgProTrpAlaAspHisArgArgArgProArgLysGln 477
QY 1332 CGACCGCGCGCAGGTGGCGCATACCGCACTCGAACAACCTACCGACGCGCGCGCG 1391
DB 478 Arg-----ArgGlnSerGlnLeuHisPro-----GlyArgArgArg 491
QY 1392 CATCATCTCCCTGATCGTCAACGCGCGCGCTGCGCAAGACGAGCGCT----- 1442
DB 492 GlnProAlaLeuGlnArgAlaArgProAlaGlnArgAlaValAlaValGlnTrpArg 511
QY 1443 ---CACCGCGCAGCA-----TGC----- 1457
DB 512 ArgHisArgProAlaArgArgArgHisGlnProLeuArgGlyCysValGlnProArg 531
QY 1458 -----CGCGCGCTGCTGCGCACGCGCGCGCTGACGCGCGCGCGCGCGCTT 1505
DB 532 GlnAlaGlnLysArgProArgArgCysArgGlnProCysArgGlnProAlaGlnProPhe 551
QY 1506 ACAATCGCGGTGAAGCGCGCGCGTGCCTCCCAACCTGAGCAACCGCTA 1565
DB 552 ArgGlyAspArgGlnArgGlnGlnArgGlnProAlaCysGlnProAspArgGlu----- 569
QY 1566 CATGCGCCCTCCCTCGCTG-----CAATAGCGCGCGCGCATCTCTGGGCAAGA 1619
DB 570 -----ProserArgArgAlaGlnLysGlnValGlyProAla-----HisArgGluAspGly 586
QY 1620 AGTGATGGCGCGCGCTGCGCGCGCAGCGCAAGCGCG----- 1658
DB 587 AlaAsp-----ArgLeuHisArgArgArgGlnArgGlnArgGlnArgGlnArgGlnArg 605
QY 1659 -----GAAATGACCTTCGACACTGTCCAGCGCGCAAGCGCGCGCTCGACCTT 1712
DB 606 HisLeuAspHisHisProAlaArgGlnAspArgGlyHisArgGlnHisArgHisGlyAla 625
QY 1713 GTCGCAAGC-----CGAGCTGGCGCGCATGCGCTT 1742
DB 626 ThrAlaGlyArgArgArgLeuArgHisArgArgArgHisProGluGlnPro 642

```

```

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21920
; LENGTH: 822
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-10-419-128-21920

Alignment Scores:
Pred. No.: 6,64e-08 Length: 822
Score: 431.50 Matches: 223
Percent Similarity: 35.848 Conservative: 34
Best Local Similarity: 31.108 Mismatches: 233
Query Match: 13.08% Indels: 227
DB: 6 Gaps: 42

US-10-009-782-1 (1-1758) x US-10-419-128-21920 (1-822)
QY 45 CGATTCGACGCGCTTCGACCTGCTGCGGCGCGCACCTCATCGACGCGACGAC 104
DB 2 ArgArgProValAspGlnProGlyProAlaArgArgHisProArgArg----- 18
QY 105 CCGGCGCGCGCGCGCGCA-----CTGGGCGTGGCGCGCGACCG-----CATCGCGC 152
DB 19 ProGlyLysArgArgArgGlyAlaProGlyGlnArgArgArgProAlaAlaAlaGlnArg 38
QY 153 CATCGGCACTGTGCGGACCGCGCGCGCACCGCGGTGCGACGTGCGGCGCTGTGT 212
DB 39 GlnArgArgGlnProArgArgArgGlnAlaGlnProGlyAlaAlaAlaLysHisAlaArg 58
QY 213 GCGCGCGGCTTCATGACGCGCACCGCACCGCA----- 245
DB 59 GlnProArgGlnLysArgGlnArgGlnArgHisProArgGlnProGlyArgHisGlnProAlaPro 78
QY 246 ---CGACACTACTCTGAG---GCGTGCAGCATGACGCGCGCGCATGCGCGCGCT 299
DB 79 GlyArgArgArgProAlaGlnProAlaGlnArgArgProAspLeuGln-----ProGlyArg 96
QY 300 CACACGCGGTGTCAGCGCGCAATGCGCGCATCAAGCTGCG-----GCCCT 344
DB 97 ArgProArgGlyAlaGlnArg-----GlnProGlyGlnProProGlyHisAlaPro 113
QY 345 GGC-----GACGCGCAACCGCGCGCGCGCTT 371
DB 114 GlyProGlyArgGlnProAlaAlaArgArgTrpArgArgAlaGlnProGlyArgProPro 133
QY 372 GACCTGCTGACGA-----AGCGCGCTCTTACCGTTT 404
DB 134 GtGlnProGlyArgGlnProArgProAlaGlnArgGlnProArgGlnArgArgArgArg 153
QY 405 CGAGCGCTTCGCGCACTGACGCGCTGGCGCGCGCGCGCGCGCGCTCAACGCGCG 464
DB 154 ArgAlaGlnGlnArgGlnGlyLeuAlaAspArgHisArgAlaValAlaArgGlnGln--- 172
QY 465 CTGTATGTGGGCACTTCAAGCTGCGCGCGCGCTGATGCGGACTTGAAGCGCGCGC 524
DB 173 -----ArgArgArgHisProGlyAlaValAlaAlaGlyAsp 183
QY 525 CACGACGAGGAATCGCGCGCAT---GCGGACCTGCGCGGAGGACCATGCGCGCGG 581
DB 184 SerArgArgAlaArgArgAlaGlnProAlaGlnProLeuGlnArgGlyAlaArgArgGln 203
QY 582 CGCATGCGGCTTCGACGCGCGCGCTTCACCGCGCGCGCGCGCGCGCGCGCAACGGA 641
DB 204 ProHisArgHis-----ArgArgLeu----- 213
QY 642 GATCATGAGGTGTGCGCGCGCTGACGCGCGCATGCGGAGATAGCCACCATGCGG 701
DB 214 -----GlyTrpArgProLeuArgGlnArgProAla 723

```



```

Db      552  ArgGlyAspArgGluArgGlnGlyLeuGlnProAlaCysGlnProAspArgGlu----- 569
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      1566  CATGACCCCTCCCTCCGCGTCG-----CAATAGCGCCCAACCCCATATCTGTGGCAAGA 1619
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      570  -----ProSerArgGlnArgGlnGlyLeuGlnValGlyProVal---HisArgGluAspGly 586
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      1620  AGTATGGGCGCGCGCTGCGCGCCGAGCGCCAGCGCCCG----- 1658
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      587  AlaAsp---ArgLeuHisArgArgGlnArgGlyLeuGlnArgGlnAlaGlnArg 605
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      1659  -----GAAATGACCCCTGCAAGACCTGTGCGACGCGCCAGCGCGCATCGCGCTCGACCCCT 1712
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      606  HisLeuAspHisHisProAlaArgGlnAlaAspArgGlyHisArgGlnGlnArgHisGlyAla 625
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      1713  GTCCAGAGC-----CGAGCTGCGCCAGATCGCCCT 1742
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      626  ThrAlaGlyArgArgArgLeuArgHisArgArgHisArgProGluGlnPro 642
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 15
US-10-366-683-32204
Sequence 32204, Application US/10366683
GENERAL INFORMATION:
APPLICANT: Rubenstein, Marc J.
APPLICANT: Nolling, Jork
APPLICANT: Delonghery, Craig
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: PAT#03-04
CURRENT APPLICATION NUMBER: US/10/366, 683
PRIOR FILING DATE: 2003-02-13
PRIOR APPLICATION NUMBER: 09/252, 991
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 32204
LENGTH: 651
TYPE: PRF
ORGANISM: Pseudomonas aeruginosa
US-10-366-683-32204

Alignment Scores:
Pred. No.: 7.5e-08 Length: 651
Score: 430.50 Matches: 208
Percent Similarity: 34.57% Conservative: 35
Best Local Similarity: 29.59% Mismatches: 247
Query Match: 13.05% Indels: 213
DB: 6 Gaps: 38

US-10-009-782-1 (1-1758) x US-10-366-683-32204 (1-651)
QY      54  GCCCTTCGACCTGCTGCTGGCGGCGACCCCTCATCGACGCAACACCCGGGCG 113
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      5  AlaGlnGluProGluArgAlaGlyArgProValProAlaArgValAlaSerHisArgGlyArg 24
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      114  GCGCGCGACCTGGCGCGCTGCGCGGAGACCGCATGCGCCCATCGCGCATGTCTGGAGCG 173
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      25  ProGlnArgGlnAlaArg---LysArgProGlnAlaArgGlnArgGlyLysPheHis 43
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      174  CGCGCGCAACACCGGCGTGCAGCTGTGCGGCTGTGCTGCGCGCGCGCTTCATCATGATC 233
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      44  ArgArgSer-----ProGlyValSerGlnProValAlaArgLeu 56
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      234  GCACACCCAGACAGACAATCACTGCT-----CAGCGCTGCGACATGACGCC---CAA 284
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      57  CysAlaProGluArgAlaArgProArgProHisHisProProAlaArgArgAlaLeu 76
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      285  GATTCGGAGGCGGTCAACACGGTGTACAGGGGCA-----TTGCGGCAT 329
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      77  ProValProArgArgHisArgProHisArgArgArgArgSerArgProAlaArgHis 96
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      330  CAGCGTGGCGCGCTGCGCGACGCCAACC----- 362
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

Db 97 ArgProGly-----GlyGlyArgArgProAlaAlaGlyArgTrpGlnGlyValCys 114
QY 363 CCGCCCCCTGGACCTGCTGGAGCGAAGCGGCTTTACCGTTTCGAGCGCTTCGCGACTA 422
    ||| ||||| |||||
Db 115 ArgThrSrpLeuProGlyGlyArg-----ArgTrpGln 125
QY 423 CCTGGAGCGCTTGG-----GGCCAGCGCGCGCGCTCAAGCGCGCTGTATGGTGG 476
    ||| ||| ||| ||| ||| ||| |||
Db 126 ProGlyAsnArgGlnGlnGlyGlnHisGlyGlyAspArgValArgArgProGlyAla 145
QY 477 CCAATTCAAGCTGGCGCGCGCTCATCGCGACTTGCAGCG----- 518
    ||| ||| ||| ||| ||| ||| |||
Db 146 -----GlyAlaArgArgGlyArgGlyProAlaGlnAlaProPheArgProArgGln 162
QY 519 -----CGCGCGCCAGCGAGGAAATCGCGCGCATCGCGGACTGCG 560
    ||| ||| ||| ||| ||| ||| |||
Db 163 ProProAlaProHisArgArgPheArgArgGlnAlaGlyArgAlaAlaSerProAla 182
QY 561 CGAGGAGC----- 569
    ||| |||
Db 183 ArgHisAlaAlaHisArgArgProGlyAlaPheLeuProGlyAspAlaAspPheLeuGly 202
QY 570 CATTGGCCAGCGGGCGCATTTGGACCGGCTTTCACCGCGCTTACCGCGCGCGCGCG 629
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 203 AsnAlaGlnArgArgArgArgHis--AspArgLeuLeuAspAlaGluProHisArg 221
QY 630 CACCCAGGAGAGATCATCATGAGTGTGCGCGCGCTGAGCGCGCATGCGCGCATCTACGC 689
    ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222 SerAlaGlnAlaGlyArgArgGlyProArgGlyPheLeuGlnProHisArgHis 241
QY 690 CACCCACATCGCGG-----CGAGCGGAGCACATCGTGGCGCGCT----- 731
    ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 AlaProHisGlyArgArgLeuArgArgGlnGlyAsnProGlyArgGlyAlaGlyLeuPro 261
QY 732 -----GAGAGAAACCTTCGCGCGCGCGCGAGCTGAGCTGCGGCTGTGATCTGCA 785
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 ValArgGlyGlyArgLeuProHis-----ArgAlaSerGlyGlnAspAla 276
QY 786 CCAAGAGTATGCG-----CCAGCCCAATTTGCGCGCGCGCGAGAGCGCT 833
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 277 ProAlaAlaHisGlyGlyHisAlaAspHisArgGlnAlaSerProAlaLeuArgArgIle 296
QY 834 GCGCGTATCGAGCGCGCATGCGCGCGCGAGAGCT----- 869
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 ArgArgArgLeuArgArgArgTrpProAlaArgHisProAspArgProGlyArgGln 316
QY 870 -----CTGCGTGGAGCGGCTATCCCTACGTCGCGCGCTCCACCATGCTCAAGCAGA 920
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 LeuArgLeuPheAlaGlyProLeuArgLeuAspArgArgProArgHisValProLeuGly 336
QY 921 CCGCGTGTGCTGGCGCGAGCGACCAT-----CATCAC 953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 GlnArgLeuLeuProArgGlnArgHisGlnArgProProLeuGlnAspGlnHisArg 356
QY 954 CTGGTGCAGAACCTTCCCGAAGCTGAGCGCGCGAGCTGATGAATCCGCGCGAGCG 1013
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 LeuGlnHisArgLeuProArgPheArgArgAla--ProGly-----AsnGlyArgHis 373
QY 1014 CGGCAATCCAGTACGAGTGTGCGCGA-----GCTGCAGCGCGCGCG 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 374 ArgGlyAspHisGlyArgArgGlyProGlnProGlyGlnGlySerAlaGlyAlaGln 393
QY 1059 CCGCATCTTCTCATGATGAGAGAACCGAGCGAGCGAGCATCTCGGCTTCGCGCGAGC 1118
    ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 AlaGlnLeuLeuArgGlnGlnArgAlaGlnArg-----HisPro--LeuProPro-- 409
QY 1119 CATGATCGGCTCCGAGCGCTGCGCGA-----CGAGAGCGCGCGCATCGCGCGCT 1169
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 410 ---AspArgArgAlaGlnProValGlyGlyAspAspArgArgAlaGlySerGlnGlnArg 428
QY 1170 GTGGGGGACCTTCCCGGGGTGCTGGGCGACTA---TGCAGCGGAGCTGGGCGCTGTCC 1226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 ValArgProProProArgGlyAsnProCysLeuGlnCysArgGlnProGlyPro----- 446

```

```

QY 1227 GCTGAGACGGCGGTATGAAAGTACCGGCTGACCGCGCGCGCTTCGGCTGGCGCG 1286
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 447 -----GlnGlnArgPro-----GlyAlaAspProGlyLys 456
QY 1287 GCGGGGAGGTGCGAGCGGGGTACTTGGCGAGCTGGTGGTGTGACCGCGGAGCGCT 1346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 ValArgHisPheValHisArgAspLeuProGlnArgGlyAspAlaAspProHisLeu 476
QY 1347 GCGCGTACCGCGACCTTTCGAACACCTTACCGAGCGCG----- 1385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 TyrArgArgGlnHisProProGlnProTrpArgHisArgAspGlyProGlnTyr 496
QY 1386 -----CGCGCG-----CATCATTCCT 1403
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 497 GlnGlyArgProGlyGlyArgArgGlyLeuProGlyArgArgGlyAlaHisProAspHis 516
QY 1404 GTAGCTCAAGCGCGCGCGCT-----CTGCGAAGAGCAGCGCTT-----CACCGG 1448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 ArgHisGlnTyrArgGlnGlyThrGlnTyrLeuAlaHisArgArgLeuGlyHisArg 536
QY 1449 CCAAGATGCGCGCGCGCTGTCGACGCGCGCGCGCTGAGCGCGCGCGCGCGCTTACA 1508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 ProGlnArgGlnGlyArgAlaGlyCysArgArgHisGlnAlaAlaProGly---Gly 555
QY 1509 ATCGCGCGT-----GAACGGCGCGCGCTGCGCGCGCGCTCCACCGCTGGA 1553
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 556 ValArgArgAlaAlaLeuGlnGlyGlnArgGlyArgArgValProGlnGlnProGly 575
QY 1554 CGCAACCGCTACATGCGCGCTCC----- 1577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 AlaArgProArgAlaAspProAlaValArgGlyThrGlyProAlaGlyLeuLeuArgPro 595
QY 1578 -----CTCGGCTCCCAATAC 1592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 GlyPheValAlaGlnHisArgLeuLeuProHisAlaGlnAspLeuLeuArgProArgAla 615
QY 1593 GCGCCG-----ACCGATATCGTGGCGGAGAGATGATGG 1628
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 616 GlyProTrpProAlaLeuLeuLeuPheCysLeuArgArgArgLeuGlyGlyAspSer 635
QY 1629 CCGCGCGCTGCGCGCGAGCGGAGCGCGGAAATGACCTGCAACCTGTGCGCAGCG 1688
    ||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 636 ArgGlyTrpHisArgArgIle-----ProHisAlaAla 646
QY 1689 CAGCGGCAAT 1697
    ||||| |||||
Db 647 HisArgHis 649

```

Search completed: May 11, 2003, 12:28:05
 Job time : 108 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 11, 2003, 12:11:00 / Search time 232 Seconds

(without alignments)
9771.041 Million cell updates/sec

Title: US-10-009-782-1
Perfect score: 3299

Sequence: 1 gaattccactgacgcgcga.....ccctgagctacgagaagcct 1758

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4569144 segs, 644733110 residues

Total number of hits satisfying chosen parameters: 9138288

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPOO/US10009782/runat.07052003-122318.23205/app.query.fasta.1.1927
-DB=Pending_Patents_AA_Main -QFMT=fastaln -SUFFIX=rapm -MINMATCH=0.1 -LOOPEXT=0
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10009782@cgn.1.1.401 -runat.07052003.122518.23205 -NCPU=6 -ICPU=3
-NO_XLPYX -NO_MAP -LARGESUBV -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMECUT=10 -THRADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

1: Pending_Patents_AA_Main:*
2: /cgn2.6/ptodata/1/paa/PCRUS_COMB.pcp:*
3: /cgn2.6/ptodata/1/paa/US06_COMB.pcp:*
4: /cgn2.6/ptodata/1/paa/US07_COMB.pcp:*
5: /cgn2.6/ptodata/1/paa/US08_COMB.pcp:*
6: /cgn2.6/ptodata/1/paa/US081_COMB.pcp:*
7: /cgn2.6/ptodata/1/paa/US082_COMB.pcp:*
8: /cgn2.6/ptodata/1/paa/US083_COMB.pcp:*
9: /cgn2.6/ptodata/1/paa/US084_COMB.pcp:*
10: /cgn2.6/ptodata/1/paa/US085_COMB.pcp:*
11: /cgn2.6/ptodata/1/paa/US086_COMB.pcp:*
12: /cgn2.6/ptodata/1/paa/US087_COMB.pcp:*
13: /cgn2.6/ptodata/1/paa/US088_COMB.pcp:*
14: /cgn2.6/ptodata/1/paa/US089_COMB.pcp:*
15: /cgn2.6/ptodata/1/paa/US090_COMB.pcp:*
16: /cgn2.6/ptodata/1/paa/US091_COMB.pcp:*
17: /cgn2.6/ptodata/1/paa/US092_COMB.pcp:*
18: /cgn2.6/ptodata/1/paa/US093_COMB.pcp:*
19: /cgn2.6/ptodata/1/paa/US094_COMB.pcp:*
20: /cgn2.6/ptodata/1/paa/US095_COMB.pcp:*
21: /cgn2.6/ptodata/1/paa/US096_COMB.pcp:*
22: /cgn2.6/ptodata/1/paa/US097_COMB.pcp:*
23: /cgn2.6/ptodata/1/paa/US098_COMB.pcp:*
24: /cgn2.6/ptodata/1/paa/US099_COMB.pcp:*
25: /cgn2.6/ptodata/1/paa/US100_COMB.pcp:*
26: /cgn2.6/ptodata/1/paa/US101_COMB.pcp:*
27: /cgn2.6/ptodata/1/paa/US102_COMB.pcp:*
28: /cgn2.6/ptodata/1/paa/US103_COMB.pcp:*
29: /cgn2.6/ptodata/1/paa/US104_COMB.pcp:*
30: /cgn2.6/ptodata/1/paa/US105_COMB.pcp:*
31: /cgn2.6/ptodata/1/paa/US106_COMB.pcp:*
32: /cgn2.6/ptodata/1/paa/US107_COMB.pcp:*
33: /cgn2.6/ptodata/1/paa/US108_COMB.pcp:*
34: /cgn2.6/ptodata/1/paa/US109_COMB.pcp:*
35: /cgn2.6/ptodata/1/paa/US110_COMB.pcp:*
36: /cgn2.6/ptodata/1/paa/US111_COMB.pcp:*
37: /cgn2.6/ptodata/1/paa/US112_COMB.pcp:*
38: /cgn2.6/ptodata/1/paa/US113_COMB.pcp:*
39: /cgn2.6/ptodata/1/paa/US114_COMB.pcp:*
40: /cgn2.6/ptodata/1/paa/US115_COMB.pcp:*
41: /cgn2.6/ptodata/1/paa/US116_COMB.pcp:*
42: /cgn2.6/ptodata/1/paa/US117_COMB.pcp:*
43: /cgn2.6/ptodata/1/paa/US118_COMB.pcp:*
44: /cgn2.6/ptodata/1/paa/US119_COMB.pcp:*
45: /cgn2.6/ptodata/1/paa/US120_COMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2511	76.1	484	24	US-10-009-782-2
2	2481	75.2	484	22	US-09-807-788-4
3	1115.5	33.8	492	18	US-09-489-039A-9018
4	612	18.6	558	21	US-09-770-51C-2
5	453	13.8	2294	16	US-09-252-991A-17231
6	434	13.2	885	16	US-09-252-991A-17231
7	433	13.1	937	16	US-09-252-991A-16129
8	432.5	13.1	663	16	US-09-252-991A-10846
9	431.5	13.1	822	16	US-09-252-991A-19843
10	430.5	13.0	651	16	US-09-252-991A-21920
11	424.5	12.9	1002	16	US-09-252-991A-32204
12	424.5	12.9	1149	16	US-09-252-991A-27980
13	423	12.8	801	16	US-09-252-991A-25557
14	421.5	12.8	863	16	US-09-252-991A-29274
15	413.5	12.5	774	16	US-09-252-991A-16789
16	410	12.4	1706	16	US-09-252-991A-1760
17	408.5	12.4	638	16	US-09-252-991A-27068
18	407.5	12.4	672	16	US-09-252-991A-16941
19	407	12.3	639	16	US-09-252-991A-20571
20	404	12.2	1053	16	US-09-252-991A-26140
21	402.5	12.2	1418	16	US-09-252-991A-23267
22	401.5	12.2	1418	16	US-09-252-991A-23267
23	399	12.1	639	16	US-09-252-991A-28453
24	398.5	12.1	724	16	US-09-252-991A-31715
25	398.5	12.1	1213	24	US-10-020-359B-95
26	395	12.0	802	16	US-09-252-991A-25050
27	394	11.9	631	16	US-09-252-991A-30909
28	394	11.9	719	16	US-09-252-991A-30464
29	392.5	11.9	782	16	US-09-252-991A-23660
30	392.5	11.9	1225	16	US-09-252-991A-25018
31	392	11.9	720	16	US-09-252-991A-21881
32	392	11.9	1228	16	US-09-252-991A-17764
33	391	11.9	937	16	US-09-252-991A-17764
34	390	11.8	757	16	US-09-252-991A-23659
35	388.5	11.8	720	16	US-09-252-991A-31915
36	388	11.8	681	16	US-09-252-991A-24567
37	386	11.7	728	16	US-09-252-991A-28169
38	386	11.7	1706	16	US-09-252-991A-31760
39	385	11.7	762	16	US-09-252-991A-29423
40	384.5	11.9	720	16	US-09-252-991A-21881
41	384	11.6	478	16	US-09-252-991A-28496
42	383.5	11.6	611	16	US-09-252-991A-32402
43	383	11.6	830	16	US-09-252-991A-17142
44	382.5	11.6	671	16	US-09-252-991A-19375
45	382.5	11.6	701	16	US-09-252-991A-24048

ALIGNMENTS

RESULT 1
US-10-009-782-2
Sequence 2, Application US/10009782
GENERAL INFORMATION:
APPLICANT: TAKEDUCHI, Ken-ichi
APPLICANT: KOIDE, Yoshinao
APPLICANT: HIROSE, Yoshiniko
APPLICANT: MORIGUCHI, Mitsunaki
APPLICANT: ISOSE, Kimiyasu
TITLE OF INVENTION: TRANSFORMED MICROORGANISM AND PROCESS FOR PRODUCING D-AMINOAC.
FILE REFERENCE: 21301USOCP
CURRENT APPLICATION NUMBER: US/10/009,782
CURRENT FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: PCT/JPO0/03932
PRIOR FILING DATE: 2000-06-15

US-10-009-782-1 (1-1758) x US-09-807-788-4 (1-484)

```

OY 34 ATGTCCCATGCGATTCAGAGCCCTTCAGACCTGCTGCGGGGCGGACCCCTCATTCAC 93
DB 1 MetAlaGluSerAspSerGlnProPheAspLeuLeuValGlyGlyThrLeuLeuAsp 20
OY 94 GGCAGCAACACCCCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 153
DB 21 GlySerAsnThrProGlyArgArgAlaAspLeuGlyValArgGlyAspArgIleAlaAla 40
OY 154 ATCGGCGATCTGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 213
DB 41 IleGlyAspLeuSerAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 60
OY 214 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 273
DB 61 AlaProGlyPheIleAspSerHisThrHisAspAspAsnThrLeuLeuArgArgAsp 80
OY 274 ATGACGCGCAAGATCTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 333
DB 81 MethTrpProIysIleSerGlnGlyValIleThrValValIleThrGlyAspGlyIleSer 100
OY 334 CTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 393
DB 101 LeuAlaProLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 120
OY 394 TCTTACGCTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 453
DB 121 SerThrArgPheGlnAlaArgPheAlaAspIleAspIleAspAlaAlaAlaAlaAla 140
OY 454 GTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 513
DB 141 ValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 160
OY 514 CAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 573
DB 161 GlnArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 180
OY 574 GCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 633
DB 181 AlaSerGlyAlaIleGlyIleSerThrGlyAlaPheThrProAlaAlaAlaAlaAla 200
OY 634 ACCGAAAGATCATCGAGGTGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 693
DB 201 ThrGluGluIleIleGluValIleCysArgProLeuSerIleIleGlyIleThrAla 220
OY 694 CACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 753
DB 221 HisMetArgAspGluArgGluHisIleValAlaAlaAlaAlaAlaAlaAlaAla 240
OY 754 CCGGAGCTGAGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 813
DB 241 ArgGlyLeuAspValAlaProValIleIleSerHisIleValMetGlyGlnProAsp 260
OY 814 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 873
DB 261 GlyArgSerArgGluThrLeuProLeuIleGluAlaAlaAlaAlaAlaAlaAlaAla 280
OY 874 CTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 933
DB 281 LeuAspAlaIleThrProValIleAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 300
OY 934 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 993
DB 301 AlaGlyArgThrIleIleThrTrpCysLysProPheProGluLeuSerIleArgAsp 320
OY 994 GATGAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1053
DB 321 AspGluValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 340
OY 1054 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1113
DB 341 AlaGlyAlaIleIleThrPheMetMetAspGluProAspValGlnArgIleLeuAlaPhe 360

```

```

OY 1114 CCGACCATGATCGGCTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1173
DB 361 ProThrMetIleGlySerAspGlyLeuProHisAspGluAlaProHisAspArgLeu 380
OY 1174 GGCACCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1233
DB 381 GlyThrPheProArgValIleGluGlyHisIleValAlaAlaAlaAlaAlaAlaAla 400
OY 1234 ACGGCGGATGGAAGATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1293
DB 401 ThrAlaValIleThrPheMetThrGlyLeuThrAlaAlaArgPheGlyLeuAlaGly 420
OY 1294 CAGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1353
DB 421 GlnLeuGlnAlaGlyThrPheAlaAspLeuValAlaPheAspProAlaThrValAla 440
OY 1354 ACCGCGACCTTCGAACACCTTCACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1413
DB 441 ThrAlaThrPheGlnHisProThrGluArgAlaAlaAlaAlaAlaAlaAlaAla 460
OY 1414 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1473
DB 461 GlyAlaProValIleProGlnGluGlnAlaAlaPheThrGlyGlnHisAlaGlyArg 480
OY 1474 CCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1485
DB 481 ArgThrAlaAla 484

```

RESULT 3

US-09-489-039A-9018
Sequence 9018, Application US/09489039A
GENERAL INFORMATION:

APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 9018
LENGTH: 492
TYPE: PRN

ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9018

Alignment Scores:

Pred. No.:	7.8e-55	Length:	492
Score:	115.50	Matches:	235
Percent Similarity:	62.708	Conservative:	71
Best Local Similarity:	48.168	Mismatches:	173
Query Match:	33.818	Indels:	9
DB:	18	Gaps:	4

US-10-009-782-1 (1-1758) x US-09-489-039A-9018 (1-492)

```

OY 19 GAAGAGAGATTCATTCATGTCACATCCGATTCGACGCTTGGACCTGCTGCTCGGGG 78
DB 9 LysGlyValIleIleAlaMetLys-----ValAspTrpLeuPheLysAsn 22
OY 79 GGCACCTTCATCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 138
DB 23 ValThrValIleAspGlySerGlyGlyProGlnTrpArgAlaAspValAlaValIleGly 42
OY 139 GACCGCGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 198
DB 43 AspArgIleMetAlaIleAlaProAlaLeuAspValAlaAlaGluGlnValIleAspGly 62
OY 199 TCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 258
DB 63 GlnGlyArgValLeuAlaProIlePheIleAspValHisThrHisAspAspIleAsnVal 82

```

QY 259 CTCAGCGCTGCGACATGACGCCCAAGATCTCGAGGGGCTGCACCGGTGCTACAGGCGC 318
 DB 83 IIEATGmetProglutinyLeuProlysLeuSerGlnGlyValThrValIleValGly 102
 QY 319 AATTGGCGCATAGCCCTGGCGCGCTGGCG---CACGCCCAACCCGCCGCCCTTCGAC 375
 DB 103 AsnGlyIleSerAlaIleThrAlaThrMetArgGlyGluValProAspProMetAsn 122
 QY 376 CTGGCGAGAGGGGGCTTACCGTTTCAGAGCGCTTCGCCGACTACGTGACCGCTTG 435
 DB 123 LeuLeuGlyGlnGlnGlnHisPheIleLeuProThrValGluAlaValHisAlaVal 142
 QY 436 CGGGCCACGCGCGCGCGCTCAACGCCCTGTATGTGGCGCATTCAGCGCTGGCGCGC 495
 DB 143 GluAlaIleArgProSerLeuAsnValGlyThrLeuIleGlyHisThrAlaLeuArgAsn 162
 QY 486 GCGGTCATCGCGGATTCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAC 555
 DB 163 AsnHisMetAspAspLeuPheArgProAlaAsnGlnThrGluIleAlaGlyMetArgVal 182
 QY 556 CTGGCGGAGGAGCGCGCGCGCGCGCGCGCGCATTCGCGCGCGCGCGCTTCACCG 615
 DB 183 GlnLeuArgAspAlaLeuArgGlnGlyAlaLeuGlyLeuSerThrGlyLeuAlaVal 202
 QY 616 CCGCAT 675
 DB 203 SerAlaPheGlnSerThrThrGlnGluValMetAlaLeuAlaGlnGluLeuAlaVal 222
 QY 676 GCGCGCATCTACG 735
 DB 223 LysGlyValIleThrThrHisLeuArgSerGluPheGlnProIleLeuGlnAlaLeuAsp 242
 QY 736 GAAACCTTCG 795
 DB 243 GluAlaPheArgIleGlyArgHisGlyAsnValProValIleSerHisIleCys 262
 QY 796 ATGGCAT 855
 DB 263 AlaGlyAlaIleAsnThrGlyArgThrLysGlnThrLeuAlaPhePheAspGluMetArg 282
 QY 856 GCG 915
 DB 283 GlnGlnGlnIleAlaCysAspCysArgProIleSerAlaSerSerThrLeuAsp 302
 QY 916 CAGGACGCGCGCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAA 975
 DB 303 MetLysGlnValThrAspGlnPheAspIleValIleThrTrpSerGluAlaGlnProGlu 322
 QY 976 CTGAGCG 1035
 DB 323 GlnAlaGlyLysThrLeuGlnGlnIleAlaAspGlnThrGlnValSerLeuHisAspAla 342
 QY 1036 GTGGCG 1095
 DB 343 AlaAlaArgLysMetProAlaGlyAlaIleThrHisAsnMetAspGlnGlnAspValArg 362
 QY 1096 CGCATCTGCG 1155
 DB 363 ArgValMetArgTrpValThrMetIleGlySerAspGlyLeuProAsnAspPromet 382
 QY 1156 CCGGATCG 1215
 DB 383 ProHisProArgLeuThrGlyAlaPheProArgValLeuGlnHisIleThrSerAlaGlyAspGlu 402
 QY 1216 GCGCTGTTCG 1275
 DB 403 GlnLeuPheProLeuThrThrAlaValHisLysMetThrGlyLeuSerAlaAlaArgPhe 422
 QY 1276 GCGCTGCG 1335
 DB 423 GlnLeuAlaAspArgGlyLeuValLysIleGlyThrPheAlaAspLeuValLeuPheAsp 442

QY 1336 CCGGCG 1395
 DB 443 ProGlnThrValArgAspValAlaSerPheSerAspProLysArgProAlaAspGly 462
 QY 1396 CATTCGCGTACGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1452
 DB 463 GluAlaValMetValAsnGlyValMetSerTrpGlySerAspLysIleThrGlyArg 482
 QY 1453 CATGCG 1476
 DB 483 ---AlaGlyArgPheLeuArgArg 489
 RESULT 4
 US-09-770-517C-2
 ; Sequence 2, Application US/09770517C
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitsubishi, Kazuya
 ; APPLICANT: Yamamoto, Hiroaki
 ; APPLICANT: Matsuyama, Akio
 ; APPLICANT: Tokuyama, Shinji
 ; TITLE OF INVENTION: D-AMINOACYLASE AND GENE ENCODING THE SAME
 ; FILE REFERENCE: 06501-072001
 ; CURRENT APPLICATION NUMBER: US/09/770,517C
 ; PRIORITY FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: JP 2000-019080
 ; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: JP 2000-150578
 ; PRIOR FILING DATE: 2000-05-22
 ; NUMBER OF SEQ ID NOS: 27
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 558
 ; TYPE: PRT
 ; ORGANISM: Hypomyces mycophilus
 US-09-770-517C-2
 Alignment Scores:
 Pred. No.: 2,626-26
 Score: 612.00
 Percent Similarity: 45.32%
 Best Local Similarity: 31.38%
 Query Match: 18.55%
 DB: 21
 Gaps: 13
 US-10-009-782-1 (1-1758) x US-09-770-517C-2 (1-558)
 QY 61 GACCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 117
 DB 4 GluIleuPheHisSerAlaThrValIleThrGlyAspGluAlaGlnProPheVal 23
 QY 118 GCGGACCTGGCG 171
 DB 24 AlaAspValLeuValSerLysGlyLeuIleAlaLysIleGlyAsnProGlySerLeuAsn 43
 QY 172 GCG 228
 DB 44 AlaThrProAspThrArgHisLysAspValThrGlyTrpIleLeuSerProGlyPheIle 63
 QY 229 GACTCGCACCCACGACGACCACTACCTGCTGCGCGCGCGCGCGCGCGCGCGCG 288
 DB 64 AspMetHisLisHisSerAspLeuTrpLeuLeuSerHisProAspHisGluAlaValIle 83
 QY 289 TCGGAGGCGTCAACGACGCGTGTGCACGCGCAATTGCGCGCATGACGCGCGCGCG 348
 DB 84 ThrGlnGlyCysThrThrGlnValIleGlyGlnAspLysIleSerTrpAlaProIleArg 103
 QY 349 CAGCGCAACCGC-----CCGCGCGCGCGCG 372
 DB 104 AsnValAspLysLeuArgAlaIleArgGlnIleAlaGlyTrpAsnGlyAsnProThr 123
 QY 373 GAC-----CTGCTGACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
 DB 124 AspGlnGlnCysArgThrThrLeuLysGlyValGlyMetPheGlnTrpGlnThrIleGly 143


```

Db 231 ProAlaAlaSerAlaTrpCysProThrCysAsnSerAlaArgProSerAlaIleAlaSer 250
QY 659 -----GGCCGCTGAGCGCGCATGGCGGATCTACGCA 691
Db 251 ThrAlaArgArgSerProSerThrAsnGlyArgProAlaAlaCysProThrThrPro 270
QY 692 CCCACATGCGCG---ACGAGGCGAGACATCTGGCGCGCGCTGAGGAAACCTTCGCA 748
Db 271 ProAlaAsnAlaProAlaSerAlaAlaAlaAsnTrpProThrPylAsnProSerPro 290
QY 749 TCGCGCGCGAGCTGGAGCTCGCGGTGATCTCGACACACAGGTCAAGGCGGCGGCGCA 808
Db 291 AlaMetProProSerThr-----296
QY 809 ATTGGCGCGCTCGCGGAGACGCTGC-----CGCTATGACGAG 847
Db 297 ---SerAlaThrArgLeuAlaAlaCysGlyHisAspGlyAspAlaGlyArgAlaSerGln 315
QY 848 CCGGCATGCGCG-----859
Db 316 ThrProThrArgAspThrSerProCysValProAsnGlyValAsnTrpArgArgGly 335
QY 860 -----GCCAGAGCTCTCGCTGGAGCGGTATCCCTACGTGGCGCGGCTCA 904
Db 336 ProThrSerAlaValAlaAlaCysThrSerProThr-----ProArgTrpProGlyMet 352
QY 905 CCATGCTCAAGCAGGACCGCGCTGC-----TGTGCGCGCGAGCGACCATATCA 952
Db 353 ProAlaGlyProAlaArgProSerCysThrAlaValAlaAlaTrp-----AlaProAlaArg 370
QY 953 CCGTGTGCAAGCCTTCCCGGACGTAGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGAGC 1012
Db 371 GlyGlyArgAlaAlaAlaAlaProArg-AsnSerCys-----SerAlaGlyArgLe 386
QY 1013 GCGGCAATCCAGTACGAGCTGTGGCGCGGAGCTGCA-----GCGGCGCG 1057
Db 386 uArgArg---ArgValaArgArgSerAlaArgAlaArgAlaTrpArgProThrProAlaGlyArg 405
QY 1058 GCGGCATCTACTATGATGAGACGACCCGACGTGCGGCGGCGGCGGCGGCGGCGGCGG 1111
Db 405 GAlaThrProAlaSerAspAlaArgArgAspSerAlaSerArgGlyArgProAlaThrAl 425
QY 1112 GCCGACCATGATGCGGTCCGAGCGGCTGCCGCGAGCAGCGCGCGCGCGCGCGCTGT 1171
Db 425 AlaAlaSerAlaProAlaAlaTrpValSerAlaAlaAlaArgArgThr---SerSerAlaProI 444
QY 1172 G---GGGACCTTCCCGCGGTGCTGGGCGACTATGCGCGCGACCT-----1214
Db 444 eAlaGlySerAlaProGlySerGlyThrAlaProArgCysHisProValaArgLysAspG 464
QY 1215 -----GGGCGTGT-----TCCGCTGAGACGCGCGGTATGGAAGATGACCGCGC 1258
Db 464 yAlaGlyProAlaSerThrGlyArgArgArgArgArgArgArgArgArgArgArgArg 484
QY 1259 -----TGACGCGCG-----1267
Db 484 GAlaArgThrGlyLeuArgArgThrAlaAlaAlaCysGlyArgProProAlaProG 504
QY 1268 -----CGCGCTTCCGCGCGGTGGCGCGCGCGCGCGG 1293
Db 504 yArgArgThrProAlaAlaArgArgArgSerAlaArgThrAlaProArgProAlaIle 524
QY 1294 CAGGTGACGCGCGGTACTTGTGCGGCGGAGCTGTGTTCGACCGCGCGCGCGCGGAT 1353
Db 524 AlaHisArgProSerAlaSerAlaThrGlyTrp-----ProGlyProAlaLeuAlaCys 542
QY 1354 ACCGCGACCTTGAACACCTTACGAGCGCGCGCGCATCTCCGTGTACGTCAAC 1413
Db 542 sProAlaAlaAlaGlyArgArgProSerProAlaProAlaProAlaProAlaProAla 562
QY 1414 GCGCGCGCGGTGTGCGACAGCAGG-----CGTTCACGCGCGGCGACGCGCGCGGTG 1467

```

```

Db 562 aArgProAlaAlaProAlaGlyArgProGlyArgSerProAlaThrAlaProAlaAla 582
QY 1468 CTCGACAGCGCGCGCGCTGAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1527
Db 582 eAlaAlaProAlaGlyArgProThrProAlaArgProPro-----AlaAlaAlaThrGlyArg 600
QY 1528 GCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1587
Db 600 gProSerAlaProProArgProAlaArgArgArgArgArgArgArgArgArgArgArg 620
QY 1588 AATA-----CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1638
Db 620 gCysSerProAlaArgProThrProArgSer-----630
QY 1639 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1698
Db 631 -AlaProGlyAlaAlaAlaThrGlyAla---ProThrThrGlyAlaAlaAlaArgPro----- 646
QY 1699 GCGGTCTGACCGCTGTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1738
Db 647 -ArgArgArgAlaAlaAlaAlaAlaProGlyLysArgArgSer 659

RESULT 9
US-09-252-991A-21920
; Sequence 21920, Application us/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rudenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMC
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21920
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21920

Alignment Scores:
Prid. No.: 4,64e-16 Length: 822
Score: 431.50 Matches: 223
Percent Similarity: 35.84 Conservative: 34
Best Local Similarity: 31.08 Mismatches: 233
Query Match: 13.08 Indels: 227
DB: 16 Gaps: 42

US-10-009-782-1 (1-1758) x US-09-252-991A-21920 (1-822)
QY 45 CGATTCCAGCCCTTGACCTGTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104
Db 2 ArgArgProValaArgGlnProGlyProAlaArgArgHisProArgArgArg 18
QY 105 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152
Db 19 ProGlyLeuArgArgArgGlyAlaArgGlyAlaProGlyLeuArgArgArgArgArgArg 38
QY 153 CATCGCGGATCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 212
Db 39 GlnArgArgGlnProArgArgGlnAlaGlyGlnProGlyAlaAlaGlyAspHisArg 58
QY 213 CCGCGCGCGGTATGACTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 245
Db 59 GlnProArgGlnArgGlyLeuArgHisProArgGlnProGlyArgHisGlnProAlaPro 78
QY 246 ---GCACAATACCTGTGTAG---GCGTGGAGCATGACGCGCGCGCGCGCGCGCGCG 299
Db 79 GlyArgArgArgProAlaGlnProAlaGlyArgProAspLeuGln-----ProGlyArg 96

```

```

QY 300 CACCGAGGTGTCACGGGCAATTGGGGCATCAAGCTGGC-----GCCGCT 344
DB 97 ArgProArgGlyAlaGlyArg-----GlnProGlyGlnProProGlyHisAlaPro 113
QY 345 GGC-----GCAGCGCAACCGCGCGCGCGCT 371
DB 114 GlyProGlyArgGlnProAlaAlaTyrArgTyrArgAlaGlyGlnProGlyArgProPro 133
QY 372 GGAGCTGCTGACGA-----AGCGGCTCTTACCGTTT 404
DB 134 GlyGlnProGlyArgGlnProArgProAlaGlnArgGlnProArgGlnArgArgArg 153
QY 405 CAGAGCGCTCGCGGACTACTGAGCGGCTGGCGGCGGCGGCGGCGGCTCAAGCGCG 464
DB 154 ArgAlaGlnGlnArgGlnGlyLeuAlaAspAlaGlyHisArgAlaValArgGlnGln 172
QY 465 CTGTATGTGGGCGCATTCACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 524
DB 173 -----ArgArgArgHisProGlyAlaValAlaAlaGlyAsp 183
QY 525 CACCGAGAGAAATCGCGGCGCAT-----CGCGGAGCTGGCGGAGAAAGCGATGGCGAGG 581
DB 184 SerArgArgAlaArgArgAlaGlnProAlaGlyProSerLeuGlyAlaGlyArgArgGln 203
QY 582 CGCGATCGGCGATTCGACGCGGCGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGAGA 641
DB 204 ProHisArgHis-----ArgArgLeu-----ArgGlnPro----- 213
QY 642 GATCATCGAGGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 701
DB 214 -----GlyTyrProGlyLeuArgGlnArgProAla 223
QY 702 CGAGAGAGCGA-----GCACATCGTGGCGCGCGCTGA 724
DB 224 GlnProArgArgProAlaPheProGlnProGlyArgGlyGlySerGlyArgGlnGly 243
QY 735 GGAAGACTTCGCGATCGCGCGCGGCGGAGCTGGA-----CGTGGCGGTGGTATCTCGACACCA 791
DB 244 ArgArgArgAlaHisArgLeuGlnProGlyArgArgAlaGlyGlnProLeuArgProAla 263
QY 792 GGGCAT-----GGCGGAGCGCAATTTCGGCGCGCTCGCGGAGAGCGCTGGCGCT 839
DB 264 GlyLeuArgArgAlaArgAlaAlaProAlaArgArg----- 275
QY 840 GATCGAGCGCGCATCGCGCGCGCGGAGAGCT-----CTCGCTGGA 878
DB 276 AspArgGlnGlnArgArgGlnProAlaArgProArgProGlnArgGlnHisAlaAlaGly 295
QY 879 CGGGTATCCCTAGCTGGCGCGCTCGACAGCT-----CAAGCAGAGCGCGCTGCT 929
DB 296 ArgTyrArgProGlnGlnArgLeuArgArgAlaGlyLeuArgGlnProGlyArgPro 315
QY 930 GCT-----GGCGGAGCGCACCATCATCATCATCGTGGCAAGCGCTT 968
DB 316 AlaThrGlyGlnProGlyGlnArgArgArgAlaHisProProHis----- 330
QY 969 CCGCGAGCTGAGCGGCGCGGCGGCTGATGAAGTCGCGC----- 1007
DB 331 ---ArgGlnTrpHisLeuArgProGlyPheArgAlaGlyAspProArgArgArgThr 349
QY 1008 -----CGAGCGCGCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1040
DB 350 AspHisGlnTrpProAlaGlyHisProAlaGlyGlnArgMetAspGlnGlnGlnAlaAla 369
QY 1041 CGAGCTGCA-----GGCGGCGCGCGCGCATCTTATGATGAGAGAGAGAGAGAGAGAGAG 1094
DB 370 SerArgThrProGlnProGlyHisArgHisLeuProProAspGlyArgGlyAlaAla 389
QY 1095 GCGGATCTCGCGGTCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1145
DB 390 GlyGlyAlaValAlaLeuHisTrpProArgArgArgLeuGlnArgArgArgProAlaGlyGln 409
QY 1146 CAGAGAGCGCGCGCATCGCGCGCTGTGGGGAGACTTCCCGGCGGTGCTGGGCGAGTATGC 1205

```

```

DB 410 ArgArgGlnLeuAlaProArgProGlyArgLeuPro-----Trp 423
QY 1206 GCGGCGCTGGCGCTGTTCGCGCTGGAGAGCGGAGTATGAGATGACCGGCTGAGCGG 1265
DB 424 GlnArgProArgHisGlnPro-----ArgArgLeuArgProGlyArg 437
QY 1266 CGCGGCTTCGCGCT----- 1286
DB 438 ArgGlnProArgProArgGlnArgArgGlnProArgArgArgArgGlnCysHisAlaAla 457
QY 1287 GCGGCGGAGCTGAGAGCGGAGA-----CTTCGCGGAGAGCTGGTGTGT 1331
DB 458 ArgArgGlnProAlaGlyGlnProTrpAlaAspHisArgArgArgArgProArgGlyGln 477
QY 1332 CGAGCGGCGGAGAGGTCGCGGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1391
DB 478 Arg-----ArgGlnSerGlnGlnLeuArgHisPro-----GlyArgArgArg 491
QY 1392 CATCGATTCGCTGTACGTACAGCGGCGGCGGCTGGCGAGAGAGAGAGAGAGAGAG 1442
DB 492 GlnProAlaAlaGlyArgAlaArgProAlaGlnArgAlaArgValAlaValGlnTrpArg 511
QY 1443 ---CACGGGCGAGCA-----TGC----- 1457
DB 512 ArgHisAspProAlaArgArgArgHisGlnProLeuArgGlyCysValGlnProArg 531
QY 1458 -----CGCGGCGGCTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1505
DB 532 GlnAlaGlyTyrArgProArgArgArgArgArgGlnProCysArgGlnProAlaGlnProPhe 551
QY 1506 ACAATCGCGGCTGACAGCGGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1565
DB 552 ArgGlyAspArgGlnArgGlnGlyLeuGlnProAlaCysGlnProAspArgGln 569
QY 1566 CATGGCGCGCGCTCGCGGCG-----CAATACGCGGCGGAGAGAGAGAGAGAGAGAG 1619
DB 570 -----ProSerArgArgAlaGlyGlyGlnValGlyProVal---HisArgGlyAspGly 586
QY 1620 AGTATGAGGCGCGGCTGGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1658
DB 587 AlaAsp---ArgLeuHisArgArgArgGlnArgGlyArgGlnArgGlnAlaGlnArg 605
QY 1659 -----GAAATGACCTCGTAACCTGTTCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1712
DB 606 HisLeuAspHisTrpAlaArgGlnAspArgGlyHisArgGlnGlnArgHisGlyAla 625
QY 1713 GTCCAGAGC-----CGAGCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1742
DB 626 ThrAlaGlyArgArgArgLeuArgHisArgArgArgHisProGlnGlnPro 642

```

```

RESULT 10
US-09-252-991A-32204
; Sequence 32204, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: MATE J. RUBENFIELD et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONA
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32204
; LENGTH: 651
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32204

```

Alignment Scores:

Pred. No.: 5, 11e-16
 Score: 430.50
 Percent Similarity: 34.57%
 Best Local Similarity: 29.59%
 Query Match: 13.05%
 Matches: 208
 Conservative: 35
 Mismatches: 247
 Indels: 213
 Gaps: 38

US-10-009-782-1 (1-1758) x US-09-252-991A-32204 (1-651)

```

QY 54 GCGCTTCGACCTGCTGCTGCGGCGGACCTTCATGAGCGAGACACACCCGGGGG 113
DB 5 AlAGInGInuProGInyArGInAGInyArGProValProInArGInyArGInyArG 24
QY 114 GCGCGCGGACCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 173
DB 25 ProGInArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 43
QY 174 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 233
DB 44 ArGInSer-----ProGInyArGInAGInAGInAGInAGInAGInAGInAGIn 56
QY 234 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 284
DB 57 CyValArProGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 76
QY 285 GATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 329
DB 77 ProValProInArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 96
QY 330 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 362
DB 97 ArGProGIn-----GInyArGInAGInAGInAGInAGInAGInAGInAGInAGIn 114
QY 363 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 422
DB 115 ArGInArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 125
QY 423 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 476
DB 126 ProGInArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 145
QY 477 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 518
DB 146 -----GInyArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 162
QY 519 -----GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 560
DB 163 ProProInArProInArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 182
QY 561 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 569
DB 183 ArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 202
QY 570 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 629
DB 203 ArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 221
QY 630 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 689
DB 222 SerAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 241
QY 690 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 731
DB 242 AlArProInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 261
QY 732 -----GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 785
DB 262 ValArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 276
QY 786 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 833
DB 277 ProAlAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 296
  
```

```

QY 834 GCGCGTTCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 869
DB 297 ArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 916
QY 870 -----CTGCGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 920
DB 317 LeuArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 936
QY 921 GCGCGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 953
DB 337 GInArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 956
QY 954 CTGCGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1013
DB 357 LeuAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 373
QY 1014 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1058
DB 374 ArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 393
QY 1059 GCGCGTTCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1118
DB 394 AlAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 409
QY 1119 CATGATTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1169
DB 410 -----ArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 428
QY 1170 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1226
DB 429 ValArGProProProProProProProProProProProProProProProProProPro 446
QY 1227 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1286
DB 447 -----GInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 456
QY 1287 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1346
DB 457 ValArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 476
QY 1347 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1385
DB 477 TyArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 496
QY 1386 -----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1403
DB 497 GInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 516
QY 1404 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1448
DB 517 ArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 536
QY 1449 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1508
DB 537 ProGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 555
QY 1509 ATCGCGCGT-----GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1553
DB 556 ValArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 575
QY 1554 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1577
DB 576 AlArTyProArGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 595
QY 1578 -----CTGCGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1592
DB 596 GInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGInAGIn 615
QY 1593 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1628
DB 616 GInProProProProProProProProProProProProProProProProProProPro 635
QY 1629 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1688
  
```



```

OY 1242 ATGAGATGACGGCTGACCGCGCGCTTGCGCTGCGCGCGCGCGAGCTGCA 1301
Db 521 -----ProserArgArg-----ArgProAlaPheAlaValGluProGln 533
OY 1302 GCGCGGCTACTGCGCGACCTGGT---GATGTCGACCGCGCGCGACGCTGCGCATCCG 1358
Db 534 ArgArgLeuAlaGlnArgProGlyAlaGlyAlaArgProAlaAlaLeuProArgArg 553
OY 1359 CACCTTCGACACCTACCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1418
Db 554 ThrProLeuArgProProAlaProAlaProGlnProValSerProAlaArgArgArg 573
OY 1419 GCGCGCTGTGCGACGACGCGCTGCAC-----CGCGCGACGA 1454
Db 574 ProGlyGlnProGlyGlyArgGlyArgGlySerLeuGlyGlyAlaAlaProArgGlyPro 593
OY 1455 TCGCGCGCGCGT-----GCTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1508
Db 594 ILevalProArgMetGlyGlyArgGlyThrAlaGlyHisArgSerArgArgProAsp 611
OY 1509 ATCCGCGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
Db 612 ---ArgArgGlyArgGlyArgProGlyArgProGlyArgPheGluArgAlaAsp 628
OY 1569 GCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 629
Db 629 -----GluArgGlnArgArgAlaGlyProGlyArgArgAlaArgProAspGlyArg 644
OY 1623 GATGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 644
Db 644 SerArgArgHisProGly---GlyAlaGlyGlyArgSerGlyAlaArgGlyThrGlyArg 663
OY 1683 GCAGCGCGCGCGG---CATCGCGCGCTGCGACCTGTCGACGCGCGCGCGCGCGCGCGCG 1739
Db 664 AlaProArgArgGlySerGlyAlaValLeuGlnArgGlyArgGlyThrGlyArgGlyArg 683

RESULT 12
US-09-252-991A-25557
; Sequence 25557, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ. ID NOS: 33142
; SEQ. ID NO 25557
; LENGTH: 1149
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25557

```

```

Alignment Scores:
Pred. No.: 1,21e-15
Score: 424.50
Percent Similarity: 34.75%
Best Local Similarity: 30.82%
Query Match: 12.87%
DB: 16
Gaps: 33

```

US-10-009-782-1 (1-1758) x US-09-252-991A-25557 (1-1149)

```

OY 15 GCGGGAAGAGAGATTTCATGTCCTCA---ATCCGATCCGACGCGCTTGACCTGCTGCT 71
Db 152 ArgAlaHisArgGlnProSerAlaProValAlaArgArgProAlaAlaAlaGlyAlaAsp 171
OY 72 GCGGCGCGACCTTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 131
Db 72 GlnProHisArgHisProLeuLeu---ArgGlyGlnHisArgPro 488

```

```

Db 172 ArgProArgAlaAspGluArgArgProGlyHisPro---ArgArgArgThrHisArgArg 190
OY 132 GCGCGGCGACCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 191
Db 191 ProArgGlnProGlnArgArgGlySerHisGlyAlaAlaArgArgAlaGlyGlnProGly 210
OY 192 CGACGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 251
Db 211 ProArgGlyArgProAsp-----His-----ProArgProArgArgArg 223
OY 252 CTACGTGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 311
Db 224 ArgProArgGlnAla-----HisHisArgGly 232
OY 312 -----CACGCGCAATTGCGCGCATGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 359
Db 233 AlaArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 252
OY 360 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416
Db 253 ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 272
OY 417 CGACTACCTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 473
Db 273 ArgThrLeuGlyGlyLeuAlaArgArgThrAlaGlyGlyGlyAlaArgArgLeuAlaGly 292
OY 474 GGG-----CATTCAGCGT-----GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 512
Db 293 AspValaSerGlnSerValProHisArgAlaAspProAlaArgHisHisArgArgArg 312
OY 513 GCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 572
Db 313 LeuGlyGlyGlyHisAlaArgArgArgArgArgArgArgArgArgArgArgArgArgArg 332
OY 573 GCGCAG---GCGCGCGCATGCGCGCATTTGACCGCGCGCGCTTGA-----CCGCGCGCGCG 623
Db 333 GlyArgValaArgLeuGlnHisHisLeuSerGlnArgLeuAlaGlyProAlaArgAla 352
OY 624 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
Db 353 AspGlyHisArgGlnGln-----358
OY 684 CTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743
Db 359 -----ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 372
OY 744 CCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 803
Db 373 GlyHisAlaGlyGluArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 390
OY 804 GCGCAATTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 863
Db 391 ProArgLeuArg-----ArgArgGlnGlnHisArgLeuPro 402
OY 864 GGA-----GTCGTGCTGACGCGCGCATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 917
Db 403 GlyAsnProGlnLeuAlaGly-----ArgGlyGlnLeuLeuHis 416
OY 918 GCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 968
Db 417 -----ArgAlaArgArgArgArgArgArgHisHisGlyArgGlyArgArgLeuGlnGlyAla 434
OY 969 -----CCCGCAATGACGCGCGCG 986
Db 435 GlnGluAlaValaArgGlnArgGlnProAspArgProLeuHisArgHisArgGlyAla 454
OY 987 CGACTGATGAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1046
Db 455 ValProGlyHisArgArgArgAlaArgArgGlnArgLeuGlnLeuArgArgGlnGlyCysArg 474
OY 1047 GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1106
Db 475 GlnProHisArgHisProLeuLeu---ArgGlyGlnHisArgPro 488

```



```

Db 391 ArgGlyLeuValLeuGlnValArgProGlyAsp 410
QY 861 CGAGG-----CGTCTCGTGGACGCTATCCCTAGCGCGGCTCCAGCAGTGC 914
Db 411 ProAlaGlyAspArgGlyAlaArgProValArgHisGlyAlaLeuArgGlnProGln 430
QY 915 GCAGGACCGGCTGCTGGCGGACGACCATCATCCCTGGTGCAGCC---CTTCC 971
Db 431 Leu-----ArgAlaGlnHisProLeuArgGlnAlaGlyLeuPro 444
QY 972 CGAGTACGCGGCGGACGCTGATGATGATGATGATGATGATGATGATGATGAT 1010
Db 445 ArgLeuAlaAlaAlaGlyLeuValLeuArgHisArgHisArgHisArgHisArg 464
QY 1011 GCGGCGCAATCCAGTACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1049
Db 465 AlaArg-----ArgAlaArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1049
QY 1050 -----GCGGCGGCGGCGGACATGATGATGATGATGATGATGATGATGATGAT 1100
Db 483 AlaGlnArgGlyAlaArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1100
QY 1101 CCTGCGCTGCGGCGGAC-----CATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1142
Db 503 ProGlyLeuHisProAspValArgHisHisProArgGlyAlaArgGlyGlyGly 1142
QY 1143 GCAGCAGCAGCGCC-----CATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1157
Db 523 LeuArgLeuAlaProAspGlnHisLeuHisProProAlaValLeuGlyArgArgPro 1157
QY 1158 -----GCATGCGGCTGCTGCGGACCTTCCCGCGCT-----GCTGGGCGCATATGC 1205
Db 543 ArgArgArgGlyHisProAlaArgHisHisProAlaArgGlyGlyGlyGlyGlyGly 1205
QY 1206 GCGGACCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1255
Db 563 HisArgPro-----ProValAlaVal-----GlnArgArgProArgGlyArg 1255
QY 1266 GCGGCGCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1319
Db 578 GlnArgArgValArgGlyGlnArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1319
QY 1320 CCGTGGTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1379
Db 598 ProArgArg-----ArgProArgHis-----ArgHisArgHisArgHisArgHis 1379
QY 1380 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1439
Db 616 ValGlnArgArgGlyAlaGlnArgGlyGlnArgGlyGlnArgGlyGlnArgGly 1439
QY 1439 -----GlnArgArgProArgGlyArg 1439
Db 636 ArgAlaGlyArgGlnArgAlaHisValGlySerHisArgArgLeuHisGlyAlaGln 1439
QY 1440 -----GTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1490
Db 656 AlaAlaAlaArgHisArgArgArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1490
QY 1491 CCG-----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1529
Db 676 SerGlnGlyArgGlyArgGly-----ArgGlyGlyGlyGlyGlyGlyGlyGlyGly 1529
QY 1530 -----GTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1577
Db 690 ArgGlyLeuArgAlaHisArgProArgHisGlnProArgArgHisGlyArgAlaAla 1577
QY 1578 CTGCGCTGCGCAATC-----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1622
Db 710 ValGlnAlaGlyGlyHisArgProArgProGlyArgArgGlyGlyGlyGlyGlyGly 1622
QY 1623 GATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1682
Db 730 AlaArgArgSerArgAlaAlaArgArgProArgProGlyGlyGlyGlyGlyGlyGly 1682

```

```

QY 1683 GCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1724
Db 750 AlaAlaArgSerAlaGlyArgLeuProProGlyGlnArgArg 763

RESULT 14
US-09-252-991A-26099
Sequence 26099, Application US/09252991A
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT FILING DATE: US/09/252, 991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26099
LENGTH: 863
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26099

```

```

Alignment Scores:
Pred. No.: 863
Score: 421.50
Percent Similarity: 32.06%
Best Local Similarity: 28.50%
Query Match: 12.78%
Gaps: 41

```

US-10-009-782-1 (1-1758) x US-09-252-991A-26099 (1-863)

```

QY 6 CCACTGATGCGGCGGAGAGAT-----TTCCATGCTCCCAATCCGATCCAGCC 56
Db 48 ProSerGlnArgArgGlyArgProProArgLeuArgLeuArgGlnArgArg 67
QY 57 CTTCGACCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 110
Db 68 IleArgProCysLeuArgSerArgArgArgCysArgHisArg-----ProGly 83
QY 111 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 164
Db 84 CysGlySerHisArgProArgGlyGlnGlnAlaProValGlnArgArgHisAlaArg 164
QY 165 GTCGACGCG-----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 209
Db 104 ArgGlyAlaValHisArgProArgLeuArgArgArgArgGlyArgArg-----AspProAla 122
QY 210 GGTGCGGCGGCGGCTTCATGACGCT-----GCACCGGCGGCGGCGGCGGCGGCGG 254
Db 123 AlaArgGlyArgGlyHisArgLeuArgProHisArgArgGlnArgArgArgArgArg 142
QY 225 CCGTGCAGCG-----TGGGCAATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 302
Db 143 ProArgHisArgArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 162
QY 303 CAGCGTGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 344
Db 163 LeuGlySerAlaGlyArgLeuArgGlnArgArgArgArgArgArgArgArgArgArg 182
QY 345 -----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 383
Db 183 ProAlaProCysGlyArgProGlyValArgArgArgProArgGlnArgProAlaLeuArg 202
QY 384 CGAGGCGGCGGCTTA-----CGTTGAGCGGCTTGC-----CGACTACT 425
Db 203 ArgHisArgLeuGlyAlaAlaArgProArgProGlyGlyGlyGlyGlyGlyGlyGly 222
QY 426 GGA-----GCGGTTGCG-----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 467

```

Db 548 ThgIyArGgIyHnIsPrOArGArGdInArGPrOgIyArGdGInPrhArGhInIa 567
 QY 1185 -----GCGGCTCGGGGCACTATGCGCGCA 1211
 Db 568 GInGInIaArGdIyArGArGhInArGdIyInIaGInArGArGdIyArGThcGdIyArG 587
 QY 1212 CCGTGGCGCTGTCCCGGTGGAGCGCGGTATGGAAAGT----- 1250
 Db 588 GInGInPrOThrIaIaIaGInIaGdIyGdIyGInIaArGdIyThrPrOaIaGInIyArG 607
 QY 1251 -----GACCGGCTGACGCGCGCGCTTGGCGTGGCG 1283
 Db 608 SerArGArGdIyPrOArGPrOgIyArGPrOThrIaIaGInIaGdIyArG 627
 QY 1284 CGGCGCGGGGAGGTGGAGCGCGGCTACTTCGCGCACTGGTGTTCGACCGCGCGAC 1343
 Db 628 GdIyIaGdIyArGdIyGdIySerArGArGdIyArGhInIaPrOaIa-----AlaHnIs 643
 QY 1344 GGTGGCGGATACGCG----- 1358
 Db 644 PrOgInIaGdIyInArGInGInIaGInIaArGdIyArGhInIaIaThrIaArGdInArPrOaI 663
 QY 1359 ---CACTTGGAGAACCC-----TACGAGCGCGCGC 1388
 Db 664 HnIsHnArGArGdIyInPrOArGhInArGPrOgIyArGSerPrOaIaArGArGdIaArGTrp 683
 QY 1389 CGGATCCATTCGCTGATCCATCAACGCGCGCGGTGGCAAGAGCGCGCTTACCGCG 1448
 Db 684 ArGdIyPrOgIyArGdIyGInIaGdIyArGdIyArGdIyArGdIyArGdIyArGhInIs 703
 QY 1449 CCAAGATGCCGCGCGCGGTCTCCACGCAAGCGCGCGCTGAGCGCGCGCGCGCTTACA 1508
 Db 704 LeuGInIaArGArGArGdIyArGdIyArG-----GdIyArGArGPrOgIyPrOaIaIaIa 721
 QY 1509 ATCCGCGGTGAAGCGCGCGCGCGCGCCGCCCTCCACCGCTGGAGCGCAACCGCTAACAT 1568
 Db 722 ThrArGArGdIaArG-----ArgIaGdIyArGdIyGInIaHnIs 734
 QY 1569 GGGCCCTCCCTCGCTCG-----CAATACGGCCCCACCGATAT----- 1607
 Db 735 GdIyPrOArGArGArGThrIaIaArGArGArGArGdIyGInIaGInArGArGdIyArG 754
 QY 1608 ---CGTGGCAAGAAATGATGGCGCGCGCGCTCGCGCGCGCGCAAGCGCGCGAAAT 1664
 Db 755 PrOPrOgIyArGdIyGdIyInIaGdIyIaGInIaIaLeuArPrOArPrOaIaPrOaIa 774
 QY 1665 -----GACCTCGCAAGACCTGTGCAAGCGCGCGCGATCGGCC 1703
 Db 775 HnIsPrOArGdIyIaGdIyGInIaGdIyInPrOgIyArGPrOaIaLeu--PrOaIaGInIaHnArG 794
 QY 1704 CTCGACCGCTGTCCAGCGCGCGAGCTGGCGCCAGATCGGCC 1741
 Db 794 InArGPrOArGPrOgInPrOaIaIaIaIaArGdIyPrO 806
 RESULT 15
 US-09-252-991A-16789
 Sequence 16789, Application US/09252991A
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 16789
 LENGTH: 774
 TYPE: PRN
 ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16789

Alignment Scores:	
Pred. No.:	4 81e-15
Score:	413.50
Percent Similarity:	33.08%
Best Local Similarity:	28.70%
Query Match:	12.53%
DB:	16
Length:	774g
Matches:	223g
Conservative:	34
Mismatches:	255
Indels:	265
Gaps:	41

US 40-005-102-1 (1-1/58) X US-09-252-991A-16789 (1-774)

QY	18	GGAGGAGAGATGTTTCATGATGCCAAATGCCAATTCGACCCGCTTGACTGCTGCTCGGG	77
Dd	58	GLYArgInAsp-----LeuProLeuLeuInProProGlySerArg	71
QY	78	CGGACACCTCTATCGACGGCAGCA-----CACCCCGGGCGGGCGGCCA	12
Dd	72	ProAspProGlyArgSerAlaArgInAlaAlaAspValAlaGluGlyProAlaGlyLysPro	91
QY	123	CCTGAGGGGTGGGGGCGACCGCATGCCCGGCATCGGCATCTGTGTGGACGCCCGCGCA	18
Dd	92	ProAlaLeuGlyGlyArgGlnHisAspGlnHisArgArgProGlnGlyProAlaArgLeu	11
QY	183	-----CACCGGGGTGCAGCTGTGGGGCGGGGTGGTGGC-----GCC	21
Dd	112	AlaGlnGlyAlaProPheArgProAlaArgAspProValProProGlyAlaSerProAspAla	13
QY	219	CGGCTTCATGACTCCCAACACCCAGCAGCAACTACTCTGTCAAGCGCTCGGACATCAG	27
Dd	132	GlyLeuHisArgGlyAlaAlaArgGlyArgArg-----	14
QY	279	GCCCAAGATCTGCAAGGGGTCAACACGGTGTCTACCGGGCAATTCGGCATCGCTGGC	33
Dd	142	-----ProGlyArgHisAlaArgGlyHisGlyGlnGlyArgAlaArgPro---	15
QY	339	GCCGCTGGCGACGCCACACCCCGCCCGCCCTGGACATGCTGTGACAGCAAGCGCTCTTA	398
Dd	157	-----AlaGlnAspArgInProAlaValAlaGlyLysPro---GlyHisArgProLeuGly	173
QY	399	CGGTTTCGAGCCTTCGCCGACTACTGGA-----CGCGTGGCGGGCCAC	443
Dd	174	AspGlyArgGlnValAlaGlnAlaGlyLeuGlyLeuArgGlyHisArgArgAspArgAspAla	193
QY	444	GCCGGCGGCGGCGACACGCCGCGCTGTATGCT-----GGGCATTCAAAGCTGGCGCGCGGT	500
Dd	194	AlaGlnArgArgGlyThrLeuArgLeuProAlaLeuGlyPro---GluArgLeuArgGlnLeu	212
QY	501	CATGCCGACTTGCACGCCGCCGCCAC---CGACGAGGAATGCGGGCCATGGCGGACCT	557
Dd	213	GlnArgGlyThrAlaGlnHisArgHisLeuProProGlyInProGlyAlaProGlyPro	232
QY	558	-----GCCGAGGAAGCCATGCGCAGCGCGGCCATCGGCATTCGACGG	602
Dd	233	HisGlyLeuAspArgGlnGlyArgGlyTyrGlnAspLeuArgLeuProArgHisProGlyArg	252
QY	603	CGCCTCTACCC-----GCCCGCGCGCGCGGCCACACCGAAGAGAT	644
Dd	253	HisArgLeuProHisHisHisAspGlnAlaArgProArgAlaGlyArgLeuGlyArgGlyArg	272
QY	645	CATCGAGGTGTGCCGCGCTGTAGCGCGCATGGCGGCATCTACGC-----	689
Dd	273	HisArgGly-----GlySerArgHisAlaAspProAlaGlyLysLeu	286
QY	690	---CACCCACATGCGCGACAGCAAGCGACACATCTGTGGCGCGCTGAGAGAACTTGG	746
Dd	287	AlaAspArgProGlyHisArgLeuGlnAlaHisArgGlnAlaAlaArgGlyHisHisArg	306
QY	747	CATCGCGCGGACCTGAGAGT-----GCCGGGTGTGATCTC	782
Dd	307	HisArgProGlyAlaAspArgHisAlaAspAlaAlaGlnGlyGlyArgGlyGlyGlyLeu	326
QY	783	-----GCACCAAGAAAGT-----	794

[illegible]

```
Db 658 GlyAspProGluGlyArgHisArgSerValProGlnGlyValArgArgGlyLeuArgArg 677
QY 1512 CGCGTGTGAACGGGGCGGCGTGCCTCCCTCC-----AACCTGGAGGC 1556
    |||||
Db 678 ArgArgGluMetAlaGlyAspProGlyAlaAlaValGlyHisLeuArgValAlaGlyArg 697
QY 1557 AACCCGCTACATGGCCCCCTCC-----CGCTCGCAATACGGCCC 1598
    |||||
Db 698 LeuHisLeuHisProAlaSerThrLeuLeuArgThrHisArgArgGlyAlaAlaGlyHis 717
QY 1599 ACCCGATATGT-----GGGCAAGAGATGATGGCGC----- 1631
    ||| |||
Db 718 ArgArgArgArgAlaGlyAlaGlyAlaGlyGlyAlaGlyArgLeuGlyAspHisArgPro 737
QY 1632 ---GCCCTCGCGCGCGAGCGCAAGCGCGAATAATGACCTGCACAGCTGTGCGCAGGC 1688
    ||| |||||
Db 738 HisLeuProCysArgGlnHisGlnGly-----ArgGlnProGlyArgProLeuProAla 755
QY 1689 CAGCGGCATCGCGCTCGACCTGTCCAAAGCGCGAGCTGGGCCAGATCGC 1739
    ::| |||||
Db 756 ArgAlaArgArgGlyThrGlnGlyLeuGlnLeuLeuArgLeuProSerArg 772
```

Search completed: May 11, 2003, 12:25:17
Job time : 262 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - protein search, using frame_plus.n2p model

Run on: May 11, 2003, 12:16:00 ; Search time 55.5 Seconds

(without alignments)
5829.953 Million cell updates/sec

Title: US-10-009-782-1

Sequence: 1 gattacacattgacgcgga.....ccctgacacagagaagcctt 1758

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 349150 segs, 92035710 residues

Total number of hits satisfying chosen parameters: 698300

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame.n2p.model -DEV=xl1p
-CGN2_1/USPTO/spool/US1000782/runat_07052003_122319_23259/app_query.fasta.1.1927
-DB=PubMed_Applications_AA -QMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-DOOPCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100
-THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USPR=US10009182 -GCCN_1.1.12 -tunat_07052003_122319_23259
-NCP=6 -ICRP=3 -NO.XLPEX -NO.MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:

1: /cgn2_6/ptodata/2/pubpa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpa/PTCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpa/US07_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpa/PTCT_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	612	18.6	558	10	US-09-770-517C-2
2	310	9.4	558	9	US-09-975-719-277
3	296.5	9.0	1367	10	US-09-801-368-108
4	289.5	8.8	5179	9	US-10-025-380-1068

Result	Score	Match	Length	DB ID	Description
5	289.5	8.8	5179	10	US-09-922-217-1068
6	289.5	8.8	5179	10	US-09-833-263-1068
7	285.5	8.7	941	12	US-10-124-557-14
8	285.5	8.7	1022	12	US-10-124-557-84
9	285.5	8.7	1038	12	US-10-124-557-74
10	285.5	8.7	1049	12	US-10-124-557-58
11	285.5	8.7	1140	12	US-10-124-557-104
12	285.5	8.7	1270	12	US-10-124-557-44
13	285.5	8.7	1313	12	US-10-124-557-42
14	285.5	8.7	1313	12	US-10-124-557-142
15	285.5	8.7	1320	12	US-10-124-557-50
16	285.5	8.7	1320	12	US-10-124-557-46
17	285.5	8.7	1320	12	US-10-124-557-60
18	285.5	8.7	1354	12	US-10-124-557-48
19	285.5	8.7	1361	12	US-10-124-557-40
20	285.5	8.7	1363	12	US-10-124-557-52
21	285.5	8.7	1404	12	US-10-124-557-52
22	285.5	8.7	1404	12	US-10-124-557-62
23	285.5	8.7	1404	12	US-10-124-557-62
24	270	8.2	1464	9	US-10-060-036-159
25	258.5	7.9	1255	9	US-09-996-069-10
26	254.5	7.7	2796	9	US-09-870-759-114
27	254.5	7.7	595	9	US-09-854-133-187
28	250	7.6	595	10	US-09-738-973-187
29	249.5	7.6	447	9	US-09-975-719-351
30	244.5	7.4	2011	9	US-10-176-847-56
31	244.5	7.4	714	9	US-10-233-885-44
32	238	7.2	362	10	US-09-866-562-61
33	237.5	7.2	550	9	US-09-976-740-47
34	237.5	7.2	550	12	US-10-023-529-47
35	237.5	7.2	550	12	US-10-023-529-47
36	237.5	7.2	2796	9	US-09-870-759-114
37	232	7.0	1806	10	US-09-919-497-56
38	228.5	6.9	399	9	US-09-764-868-850
39	226.5	7.0	447	9	US-09-975-719-351
40	225	6.9	595	9	US-09-854-133-187
41	225	6.9	595	10	US-09-738-973-187
42	223.5	6.8	699	9	US-10-121-988-143
43	222	6.7	1023	9	US-09-893-519A-14
44	221.5	6.8	558	9	US-09-975-719-277
45	221.5	6.7	1336	9	US-09-945-901-58

ALIGNMENTS

RESULT 1
US-09-770-517C-2
Sequence 2, Application US/09770517C
Patent No. US20020151035A1
GENERAL INFORMATION:
APPLICANT: Mitsuhashi, Kazuya
APPLICANT: Yamamoto, Hiroaki
APPLICANT: Matsuyama, Akiohbu
APPLICANT: Tokuyama, Shinji
TITLE OF INVENTION: D-AMINOACYLASE AND GENE ENCODING THE SAME
FILE REFERENCE: 06501-072001
CURRENT APPLICATION NUMBER: US/09/770,517C
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: JP 2000-019080
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: JP 2000-150578
PRIOR FILING DATE: 2000-05-22
NUMBER OF SEQ ID NOS: 27
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 558
TYPE: PRT
ORGANISM: Hypomyces mycophilus
US-09-770-517C-2
Alignment Scores:
Pred. No.: 2.4e-20
Score: 612.00
Length: 558
Matches: 171

```

01 GACCTCTCTCTGGCGGGCGGACCCCTCATCGACGCGACCAACACCCCGGGCGG---CGC 117
Db 4 GlnIleuPheHisSerIalThrValIleThrGlyArgSerIuAlaIleGlnProPheVal 23
0Y 118 GCGGCACTGGGGGTGGCGGGCGGACCGCATCGCGCCATCGGCGATCTGTGCG---GAC 171
Db 24 AlaAspValIleuValSerItyrGlyLeuIleAlaIlyIleGlyAsnProGlySerIleAsn 43
0Y 172 GCGCGCGCGACACCGCG---GTGACGCTGTGCGGGCTGTGGTGGCGGGCGGGCTTGATC 228
Db 44 AlaIthrProAspThrArgIleuAspAlaIleThrGlyIleuLeuSerProGlyPheIle 63
0Y 229 GACTGCGACACCCAGAGACGACACACTACTGCTCAAGGCTGGCGACATGACGCCCAAGATC 288
Db 64 AspMetHisAlaHisSerAspLeuIlyrIleuSerHisProAspHisGlnAlaIlyIle 83
0Y 289 TCGCAGGGGCGTCAACGAGGTGGTGTCAAGGGCAATGGCGCATACCGCGGGCGGGTGGCG 348
Db 84 ThrGlnGlyCysThrThrGlnValValGlyIleAspIlyIleSerIlyIleThrAlaProIleArg 103
0Y 349 CACGCCACCGCG-----CGCGCGCGCGCTG 372
Db 104 AsnValAspGlnIleuArgAlaIleArgGlnGlnIleAlaGlyThrProAsnGlyAsnProIle 123
0Y 373 GAC-----CTGCTGGACGACAGCGCGCTTACGCTTTCAGCGCTTGGCC 417
Db 124 AspGlnGlyCysArgThrThrIleuIySerIyAlaIlyIleThrPheGlnIlyrGlnThrIleGly 143
0Y 418 GACTACTCGACGCGCTTGGCGGGCGACCGCGCGCGCGCTCAACGCGCGCGCTATGAGGGCG 477
Db 144 GlnTyIleuAspCysIleuGlnIuArgAsnArgThrIalThrAsnIyAlaIleMetIleValPro 163
0Y 478 CATTCACGCGCGCGCGCGCGCTATCGCGGACTTGCAGCGCGCGCGCGACCGACGAGAA 537
Db 164 GlnGlyAsnIleuArgIleuAlaIlySerGlyProTyArgAspThrProAlaSerIalGlnIu 183
0Y 538 ATCGCGGCGCATCGGAGCACTGGCGGAGGAAAGCCATGGACGAGCGCGCGCATTTGG 597
Db 184 IlegIAspGlnIleGlnIleuAlaArgIuAlaMetIalGlnIyAlaValGlyMetSer 203
0Y 598 ACCGGCGCTTCTACCGCGCGCGCGCGCGCGACACCGAAGATCATCGAGGTGCG 657
Db 204 SerGlyLeuThrTyIThrProGlyMetTyIAlaSerThrSerGluIleuAlaSerIleCys 223
0Y 658 CGCGCGCGGAGCGGCGAT-----GGCGGCGATTCAGCCACCCACATCGCGCAGAAAGC 711
Db 224 AlaIalAlaIleuAlaGlnIuPheProGlyAlaPheTyIAlaProIHisIAspSerTyGly 243
0Y 712 GAGCACATCGTGGCGCGGTGAGGAAACCTTCGCGATCGCGCGCGAGCGAGTGGACGGCG 771
Db 244 PheGlnAlaIleGlySerTyIArgIleuMetIleuAspIleuGlyIleSerThrGlyIySerPro 263
0Y 772 GTGGTGAATTCGACCAACAAGTCATGGGCGACCCCAATTTGGCGCGCTGGCGGAGAGC 831
Db 264 IleHisIleuThrHisAlaThrIleuAsnPheSerGlnAsnIyGlyIyAlaProValIleu 283
0Y 832 CTGCGCGGTACGACGAGCGCGCATGGCGCGC---CAGAGAGTGTCCGTGACCGCGATACCC 888
Db 284 IleserMetValAspIySerIleuAlaIleGlyValAspValThrIleuAspThrTyIPro 303
0Y 889 TACGTGGCGCGCTCCACCATGCTC----- 912
Db 304 TyIleuProGlyCysThrThrIleuAlaAlaIleuProSerThrAlaSerIalGlyIy 323
0Y 913 -----MACGAGACCGCGTGTGCTGGCC 936

```

Alignment Scores:	
Pred. No.:	6.66e-07
Score:	310.00
	Length:
	Matches:
	558
	.182

Percent Similarity: 35.07% Conservative: 27
 Best Local Similarity: 30.54% Mismatches: 221
 Query Match: 9,408 Indels: 167
 DB: 9 Gaps: 34

US-10-009-782-1 (1-1758) x US-09-975-719-277 (1-558)

```

OY 60 CGAGCTGCTGCTCCGGGGGCGGACCTTCAT-----CGAGGCGACCAACACCC 107
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34 ArgProAspArgGlyArgGlnProHisProHisGlyAspArgGlnGlnGlyGly 53
OY 108 GGGGCGGCGCGGCGGCTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 167
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54 ThrGlySerSerHisProSerProAspProAlaGlnAspArgGlnGlnGlyGly 73
OY 168 GAGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 218
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 74 GlyArgArgArgSerHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 92
OY 219 CGGCTTCATGAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 272
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93 AlaArgAlaArgGlnArgArgProAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 112
OY 273 -----CATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 320
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 113 GlyLeuGlyAlaAspProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 132
OY 321 TTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 AlaGlnGlnArgProArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 152
OY 381 -----GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 425
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153 ThrGlnProAlaAspArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 166
OY 426 GAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 485
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 167 ---ProValAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 184
OY 486 GCTGCGCGG-----CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 524
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185 ValAspArgGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 204
OY 525 CACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 584
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 GlyAlaGlnAlaAlaSerHisSerHisArgProGlnGlnGlnGlnGlnGln 222
OY 585 CATGCGCATTTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 644
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 223 -----ValGlnArgArgGlnProPheAlaArgArgProAlaArgSerGlnProAla 239
OY 645 CATGAGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 704
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 240 GlnArgLeuValArgProLeuPro----- 246
OY 705 CGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 746
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 247 -----ArgProAspArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 264
OY 747 CATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 806
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 265 LeuArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 283
OY 807 CAATTTCGCGCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 839
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 LeuGlnGlnProProAspArgArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 303
OY 840 ---GATGAGCG-----CGCGCATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 878
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 ProAspArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 321
OY 879 CGCGATTCCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 938
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 322 GlyGlnGlyProSerGlnGlyGlnGlnHisSerProGlnAlaArg-----GlyArg 339
OY 939 ACGCAACCATATCATCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 998
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 ThrPro-----AspArgGlnSerArgGlnGlnGlnGlnGlnGlnGlnGln 351
OY 999 AGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1055
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 352 ProGlnHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 371
OY 1056 CGGCGGCG-----CATCTATCTATGATGAGCGGCGGCGGCGGCGGCGGCGG 1103
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 ArgArgProArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 391
OY 1104 GCGGCT---CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 GlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 411
OY 1125 CGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1184
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 412 ArgArgHisArgArgArgProAlaArgSerProAlaGlnAlaPro----- 426
OY 1185 GCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1244
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 427 -----GlnArgAlaArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 441
OY 1245 GAGATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1298
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 442 -----GlnAlaAspArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 458
OY 1299 GAGCGGCGG-----GTACTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1349
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 459 AlaGlnHisProAlaLeuAspArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 475
OY 1350 CGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1394
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 -----ArgGlnLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 493
OY 1395 CCATTCCTGT-ACGTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1453
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 ArgLeuArgLeuThrSer-----SerGlyAlaAspArgGlnSerProAlaSer 509
OY 1454 ATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1513
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 510 -----AlaGlyLeuPheHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 517
OY 1514 GCGTAAAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1573
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 518 -----ValGlyAlaArgGln----- 522
OY 1574 CTCCCTCCGCTCGCAATACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1621
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 523 LeuProHisAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 538

```

RESULT 3
 US-09-801-368-108
 ; Sequence 108, Application US/09801368
 ; Patent No. US20020128250A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Busby, Robert
 ; APPLICANT: Call, Brian
 ; APPLICANT: Hecht, Peter
 ; APPLICANT: Holtzman, Doug
 ; APPLICANT: Madden, Kevin
 ; APPLICANT: Maxon, Mary
 ; APPLICANT: Milne, Todd
 ; APPLICANT: No. US20020128250A1man, Thea
 ; APPLICANT: Royer, John
 ; APPLICANT: Salama, Sofie
 ; APPLICANT: Sherman, Amir
 ; APPLICANT: Silva, Jeff
 ; APPLICANT: Summers, Eric
 ; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fung;


```

OY 701 GCAGCGAGAGGCGAGACATCGTGGCCGGCTGAGAGAACTTCGCGATCGGCGCGAGC 760
Db 545 SerThrThrIuSerSerSerAlaProValProThrProSerSerSerThrThrIuSer 760
OY 761 TGAGCGTCCGGGTGGTGAATCTCGACACACAGAGTATGGCGAGCCAAATTGGCCGCT 820
Db 565 SerSerThrProValThrIuSerSerThrThrIuSerSerSerAlaProValProThr 820
OY 821 CGCGCGAGAGCGTCCGCGTGAATCGAGAGCCCGCATGGCGCGAGACGTCGCTGAGAGC 880
Db 585 SerSerSerThrThrIuSerSerSerSerAlaProValProThrProSerSerSerThr 604
OY 881 CGATCCCTACGCGCGCGCTCCACCATGCTCAAGACAGAGAGCGCGTCTGCTGCGCGAGC 940
Db 605 GluSerSerSerAlaProValProThrProSerSerSerThrThr-----Glu 620
OY 941 GCACATCATCACTACCT-----GGTGCAGACCCCTTCCCGAGAC 976
Db 621 SerSerSerAlaProValThrSerSerThrThrIuSerSerSerSerAlaProValProThr 640
OY 977 TGAGCGGGGGCGCGCATGATGAAGTGGCGGGCGAGCCGGCAATTCAGTACGAGTGG 1036
Db 641 ProSerSerSerThrThrIuSerSerSerSerAlaProValProThrProSerSerSerThr 660
OY 1037 TGCGCGAGCTGACGCGCGCGCTCACTCACTCATGATGAGAGCAACCGACGCTGACAC 1096
Db 661 ThrGluSerSerSerAlaProValProThrProSerSerSerThrThrIuSerSerThr 680
OY 1097 GCATCTCTGGCGCTGGCGC-----CGAACATGATCGGCTCCGACGGCGCTCGCGACGAGC 1156
Db 681 AlaProValThrSerSerThrThrIuSerSerSerSerAlaProValThrSerSerThrThr 700
OY 1151 AGCGCGCGGATCCGCGCTGTGGGGACACTCCCGGGGGTGTGGGGGCACTTGGCGCGC 1216
Db 701 GluSer-SerSerAlaProValProThrProSerSerSerThrThrIuSerSerSerAl 720
OY 1211 ACCTGGCGCTGTCCCGCTGAGAGAGCGCGGTATGAAATGACCGCGCTGACCGCGCGCGC 1270
Db 720 AlaProValProThrProSerSerSerThrThrIuSerSerSerSerAlaPro--ValProThr 739
OY 1271 GCTTGGCGCTGCGCGCGCGCGGAGCTGACGCGCGGGATTCGCCGACCTGGTGTGT 1330
Db 740 ProSerSerSerThrThrIuSerSerSerSerAlaProValThrSerSerThrThrIuSer 759
OY 1331 TCGACCGCGGCGACGCTGGCGCGATACCGCCACTTCGACACCCCTACGAGCGCGCGCG 1390
Db 760 SerSerAlaPro--ValProThrProSerSerSerThrThrIuSerSerSerAlaPro 778
OY 1391 GCATCCATTCGGTACGTCAAGAGCGCGCGCGCTGCGACAGAGCAGCGGTTCACCGCGC 1450
Db 779 ValProThrProSerSerSerThrThrIuSerSerSerSerAlaProValProThrProSer 798
OY 1451 AGCATGCGCGCGCGTGTCTGCGACAGCAGCGCGCGCTGAGCCCGCGCGACCGCTTACAAT 1510
Db 799 SerSerThrThrIuSerSerSerAlaProValProThrProSerSerSer--Asn 817
OY 1511 CGCGCGCGAGAGCGCGCGCGTGGCGCGCGCTCCCAACCTGAGAGCAACCGCTACATGG 1570
Db 818 IleThrSerSerAlaProSerSerSerThrProPheSerSerSerThrIuSerSerSerVal 837
OY 1571 CCCCTCCCT 1579
Db 838 ProValPro 840

```

Db	1501	SerProProMetThrThnProLleThrPro-----Pro	1511
QY	602	GGGCTTACCAGCCCGCCGCGCCGACACCGAAGATCATCGAGGTGTGCCGC	661
Db	1512	AlaSerThrThrThnLeuProProThrThnPro-----	1533
QY	662	CGCTGAGGGCGATGGCGGCAATCAAGCCACCCACATCGCGAGAGGCGAGCATCG	721
Db	1524	-----SerProProThrThnThr-----ThrThr	1532
QY	722	TGGCCGCGCTGAGAGAAACCTCCGCATCGGCGCCGAGACTGAGCTGGCGGTGTGATCT	781
Db	1533	ProProProThrThnThrProSerProProThr-----	1543
QY	782	CGACACCAAGAGTATGGGCCAGCCCAATTTGGCCGCTGCGGAGACGCTGCCGCTGA	841
Db	1544	-----ThrThnProLleThr-----	1548
QY	842	TCGAGGCCCGCATGCGCGCGCAGACGCTTCGCTGAGAGCGCATCCCT---ACGTGGCG	898
Db	1549	-----ProPro-----ThrSerThrThrLeuProProThrThnPro	1562
QY	899	GCTCCACCATGCTCTAACAGAGACCGCGCTGCTGCTGCGGAGACGACCATCATCACTGTG	958
Db	1563	SerProProProThrThnThrThnThr-----ProProProThr	1575
QY	959	GCAAGCCCTTCCCGCACTGAGCGGGCGGACCGATGATGAAGTGGGGCCGAGCCGGCA	1018
Db	1576	ThrThrProSerPro-----ProThrThnThr	1584
QY	1019	AATCCAAGT-----ACAGCGTGTGCCGAGCTGCAGACCGCGCGCGCA	1063
Db	1585	ThrProSerProProThrLleThrThnThrProProProThrThnThrProSerPro	1604
QY	1064	TCTACTTCATGATGAGAGAAC-----CCGAGCTGAGAGCGCATCCTGGCGTGTGGCCCA	1117
Db	1605	ProThrThrThrThrThrThrProProProThrThnThrProSerProProThrThnThr	1624
QY	1118	CGATGATGGCTCCGACAGCGCTCGCGCAGACAGAGCGCCGATCCGCGCTGTGGGCA	1177
Db	1625	ProLleThrProProThrSer-----ThrThr-----	1633
QY	1178	CCYTCCCGGGGCTGTGGGCACTATGSCGCGGACCTGGGCGCTGCCGTGAGAGCG	1237
Db	1634	-----ThrLeuProProThrThnThr-----	1640
QY	1238	CGGTATGAGATGATACCGGCGCTGCAGCGCGCGCTTCGCGCTGGCGGGCGCGGAC	1297
Db	1641	-----ProSerProProProThrThnThrPro-----	1651
QY	1298	TGCAGCGGGGTACTTCGCGCGACATGCTGTGTGACCCGCGCACGCTGGCCGATA--	1354
Db	1652	-----ProProThrThnThrProSerProProThrThnThrProSerProProLleThr	1670
QY	1355	-----CCGCACTCTTGCAACCCGACCGAGCGGCGCG-----	1387
Db	1671	ThrThrThrThrProProProThrThrThrProSerSerProLleThrThnProSer	1690
QY	1388	-----CCGCACTCATTCCTGT	1405
Db	1691	ProProThrThrThnMetThrThnProSerProThrThnThrProSer-----ProLle	1709
QY	1406	ACGTCAACGAGCGCGCGGTGTGGCAAGAGCAGCGCTACACGCGCATCGCGCGCG	1465
Db	1710	ThrThrThrThrThrProSer-----SerThrThrThrProSerProProProThrThn	1727
QY	1466	TGCTCGCAGCAGCGCGCTGAGCGCGCGCCAGCCCTTCAATCCGGCGCTGACAGGGG	1525
Db	1728	MetThr-----ThnProSerProThrThrThrProSerProProThrThnMet	1744
QY	1526	CGGGGTGCGCGCCCTCCCAACCTGAGACGCAACCGCTACATGAGCCCTCCCGCTC	1585
Db	1745	ThrThrLeuProProThrThnThrSerSerProLeuThrThnThrProLeuProPro--	1763

Db 1710 ThrrThrrThrrThrrProser-----SerThrrThrrThrrProserProProThrrThrr 1727
 QY 1466 TCGTGCAGCAGCAGCGCGCTGAGCCCGCGCCAGCCCTTACATCCGCGGAGAGGG 1525
 Db 1728 MetThrr-----ThrrProserProThrrThrrThrrProserProThrrThrrMet 1744
 QY 1526 CGCGCTGCGCGCGCGCTCCCAACCTTGAGCGCAACCGGTACATGCGCCCTCCCTCCGCTC 1585
 Db 1745 ThrrThrrLeuProThrrThrrThrrThrrSerProLeuThrrThrrThrrProLeuPro 1763
 QY 1586 GCATACGCGCGCGCGCGCGATATCG 1609
 Db 1764 SerThrrThrrProThrrThrrPhaser 1771

RESULT 6

US-09-833-263-1068
 ; Sequence 1068, Application US/09833263
 ; Patent No. US20020110547A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Clapper, Jonathan D.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Meagher, Madeline J.
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
 ; FILE REFERENCE: 210121.471C12
 ; CURRENT APPLICATION NUMBER: US/09/833,263
 ; CURRENT FILING DATE: 2001-04-10
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1068
 ; LENGTH: 5179
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-833-263-1068

Alignment Scores:

Pred. No.: 4.4e-06 Length: 5179
 Score: 289.50 Matches: 146
 Percent Similarity: 32.92% Conservative: 41
 Best Local Similarity: 25.70% Mismatches: 164
 Query Match: 8.78% Indels: 217
 Db: 10 Gaps: 29

US-10-009-782-1 (1-1758) x US-09-833-263-1068 (1-5179)

QY 20 AAGAGAGATTCATGTCACATCCGAT-----CCACGCCCTTCG 61
 Db 1383 LysIleArgValAlaAsnCysScySTRProMetAspLysCyAlaThrrThrrProserProPro 1402
 QY 62 ACCGTGCTGCTCGCGCGCGCGCGCGCTCATGAGCGAGCAACCGCGCGCGCGCGCGCG 121
 Db 1403 ThrrThrrThrrProserProProThrrThrrThrrThrrLeuPro-----Pro 1418
 QY 122 ACCGTGCGCTGCG 181
 Db 1419 Thrr-----ThrrThrrProserProProThrrThrrThrrThrrThrrProPro 1434
 QY 182 ACACCGGCTGACGCTGTCGGGCGCTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 241
 Db 1435 ThrrThrrThrrPro-----SerProThrrThrrThrrThrrThrrThrrThrr 1448
 QY 242 ACAGCAGACACTACCTGCTCAGGCGTGCAGACAGCGCCCAAGATTCGCGAGGCGTCA 301
 Db 1449 LeuProThrrThrrThrr-----Proser 1455
 QY 302 CCACGGTGTGACGGGGAATGGCGCATACAGCTGGCGCGCGCGCGCGCGCGCGCGCGCG 361
 Db 1456 ProThrrThrr-----ThrrThrrThrr 1463
 QY 362 CG 421

Db 1464 ProProThrrThrr-----ThrrProserProPro 1473
 QY 422 ACCGTGAGCGCTTGGGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 481
 Db 1474 ThrrThrrThrrProserProProThrrThrrThrrThrrThrrProserProThrr 1489
 QY 482 CAAGCTGCGCGCGCGCGCGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 541
 Db 1490 -----ThrrThrrThrrProProThrrThrrThrrThrrThrrPro 1500
 QY 542 CGCGCATGCG 601
 Db 1501 SerProThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1511
 QY 602 CGCGCTTACCG 661
 Db 1512 AlaSerThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1523
 QY 662 CGCTGAGCGCGCATGGCGCGCATACGCGCACCGACATGCGCGCGAGAGGCGACATCG 721
 Db 1524 -----SerProThrrThrrThrrThrr-----ThrrThrr 1532
 QY 722 TGGCGCGCGTGGAGGAACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
 Db 1533 ProProThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1543
 QY 782 CGACACCAAGATCATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841
 Db 1544 -----ThrrThrrThrrThrr-----ThrrThrrThrrThrr 1548
 QY 842 TCGAGCG 898
 Db 1549 -----ProPro-----ThrrThrrThrrThrrThrrThrrThrrThrrThrr 1562
 QY 899 GCTCAACATGCTCAGACGAGAGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCG 958
 Db 1563 SerProThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1575
 QY 959 GCAGCGCTTCCCG 1018
 Db 1576 ThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1584
 QY 1019 AATCCAGT-----ACGAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1063
 Db 1585 ThrrProserProThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1604
 QY 1064 TCTACTTCATGATGAGAGAC-----CCGAGCTGACGCGATCTGCGCGCGCGCGCG 1117
 Db 1605 ProThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1624
 QY 1118 CCATATGCGCTCGACGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1177
 Db 1625 ProThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1633
 QY 1178 COTTCGCGCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1237
 Db 1634 -----ThrrLeuProThrrThrrThrr----- 1640
 QY 1238 CGTATGGAATGACG 1297
 Db 1641 -----ProserProProThrrThrrThrrThrrThrrThrrThrrThrrThrr 1651
 QY 1298 TGCAGCGCGGCTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1354
 Db 1652 -----ProProThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1670
 QY 1355 -----CCGCGCGCTTGGAGAACCTTACGAGCGCGCGCG----- 1387
 Db 1671 ThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1690
 QY 1388 -----CCG 1405
 Db 1691 ProProThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 1709

code: 285.50

...ProAlaProThrAlaProLysGluProAlaProThrThrPro 414

```

Db 415 LysgluThralaProthrProLysLysLeuThrProthrThrProgluLysLeuAla 434
QY 1031 ACSTGGTCCCGAGCTCCAGCCGGCCGCACTACTATCATATGAGACACCCGAGC 1090
Db 435 ProthrThrPro-----GluLysProAlaProthrThrProgluLysLeuAlaProthr 452
QY 1091 TGCAGCGCATCTGCGCTTGGCCGCCGCACTGATGCG--GCTCCGAGCGGCTGCGGACG 1147
Db 453 ThrProgluLupProthrProthrThrProgluLupProAlaProthrThr----- 469
QY 1148 ACAGAGGCCCGCATCCGCGCTGTGGGGACCTTCCCGGGGTGGGGGACATGATGCGC-1207
Db 469 ----- 469
QY 1208 GCGACCTGGGCTGTTCCTGAGAGACGGCGGTATGGAATGACCGGCTGACCCCGC 1267
Db 470 -----Pro 470
QY 1268 CCGGCTGGGCTGAGCCGGCGGCGGAGCTGAGCGCGGCTGCTGCGGACCTGCTG 1327
Db 471 LysAlaAlaAla---ProAsnThrProLysgluProAlaProthrThrProLysgluPro 489
QY 1328 TGTTCACCCCGGCGAGCGGTGGCCGATACG---CCACCTTGAACACCTACCGAGCGC 1384
Db 490 AlaProthrThrProLysgluProAlaProthrThrProLysgluThralaProthrThr 509
QY 1385 CCGCGCGCATCATCTGCTGATGATCAAGCGCGCGCGCTGCTGCAAGACGAGCGCTCA 1444
Db 510 ProLysgluThralaPro---ThrThrLeuLysgluProAlaProthrThrProLysLys 528
QY 1445 CCGCGCGCATCGCGCGCGCTGCTGCAAGCGCGCGCGCTGCTGCAAGACGAGCGCT 1504
Db 529 ProAla-----ProLysgluLeuAlaProthrThrThrLysgluProthrThrSer 545
QY 1505 TACATCCGCGCTGAACGGGGCGGCGCGCGCTGCGCGCGCTGCGCGCGCTGAGCAACCGCT 1564
Db 546 -----ThrSerAspLysProAlaProthrThrThrProLysgluThralaPro 560
QY 1565 ACATGCGCCCTCCCTCCGCTGCGCAATAGAGCGCCGACCCGATATCGGTGGGCAAGAAATGA 1624
Db 561 ThrThrProLysgluProAlaProthrThrPro----- 571
QY 1625 TGGGCGCGCGCTGCGCGCGCGAGCGCGCAAGCGCGAAATGATGACCTGCAAGACCTGCG 1684
Db 572 -----LysgluProAlaProthrThrProLysgluThralaPro---ThrThrLeuLys 588
QY 1685 AGCGGAGCG---GCAATCGCGGTCTGACCTCTCCAGG-----CCGAGCTGGGGC 1732
Db 589 GluProAlaProthrThrProLysLysProAlaProLysgluLeuAlaProthrThrThr 608
QY 1733 AGATCGCCCTGAGTACGAGGAGC 1756
Db 609 LysgluProthrSerThrThrSer 616

```

RESULT 8
US-10-124-557-84
Sequence 84, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 84:

SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Alignment Scores:

Pred. No.: 7,75e-06 Length: 1022
Score: 285.50 Matches: 137
Percent Similarity: 33.50% Conservative: 60
Best Local Similarity: 23.30% Mismatches: 240
Query Match: 8.65% Indels: 151
DB: 12 Gaps: 24

US-10-009-782-1 (1-1758) x US-10-124-557-84 (1-1022)

```

QY 32 CCATGTCCTCAATCCGATTCACGCTTGCAGCTGCTGCGGCGGCGACCTCATCG 91
Db 248 ProthrProthrThrThrLysSerAlaProthrThrProLysgluProAlaProthrThr 267
QY 92 ACGGCGACA-----ACACCCCGGGGGCGCGCGCGACCTGGGCGCTGCGGCGAC 142
Db 268 ThrLysSerAlaProthrThrProLysgluProAlaProthr-----Thr 282
QY 143 GCATGCGCGCATCGGCGCATCTGTGAGACCGCGCGCGACACCGGGGTGACGTGTGG 202
Db 283 ThrLysgluProAlaPro-----ThrThrProLysgluProAlaProthr----- 297
QY 203 GCGTGTGTGCGCGCGCGCTTCATGACGATCGACACCGACGACGACGATGCTGCA 262
Db 298 -----ThrThrLysgluProAlaProthrThrThrLysSer 309
QY 263 GCGGTGCGCATGACGCCGCAAGATCTGCGAGCGGCGTCAACGCGTGTCAAGGCAATT 322
Db 310 AlaProthrThr-----ProLysgluProAlaPro----- 319
QY 323 GCGGATCATGCTGCGCGCGGTGCGGCGACGCAACCGCGCGCGCGCTGACCTGCTGG 382
Db 320 -----ThrThrProLysLysProAlaProthrThrThrProLys 331
QY 383 ACGAAGCGGCTTACCGCTTTCGAGCGCTTCCGCGCATCTGAGACGCTTGGCGGCA 442
Db 332 GluProAlaProthrThrProLysgluProthrProthrThrProLysgluProAlaPro 351
QY 443 CCGCGGCGG-----CGTCAACGCGCGCTGTATGATGGCGCATTCACGCTGCGCG 493

```

```

Db 352 ThrThrLysGluProAlaProThrThrPro----- 361
QY 494 CCGCGGTCATGCGGAGCTTGACGCGCGCCAGCAGAGAAATCGCGGCATCGCGG 553
Db 362 -----LysGluProAlaProThrAlaProLysGluProAlaPro 374
QY 554 ACCGCGCGAGAGAACCATGGCCAGCGCGCCATTCGACGGCGCTTCAC 613
Db 375 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 392
QY 614 CGCCCGCGCGCGCGCA---CAACGAGAGATCATGAGGTGCGCGCGCTGAGCG 670
Db 393 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrLysSerAla 412
QY 671 CGCATGGCGGCTATACGCCACCCATGCGCGAGAGGAGGACATGTCGCGCGC 730
Db 413 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 432
QY 731 TGGAGGAACCTTCGCGCATGGCGCGAGCTGACGTGCGGTGATCTCGCACACA 790
Db 433 GluProSerProThr-----ThrThrLysGluProAlaProThrThr 446
QY 791 AGGTGATGGCGCAGCCCATTTTCGGCGCTCGCGGAGACGCTGCGCTGATCGAGCGG 850
Db 447 ProLysGluProAlaPro-----ThrThrPro 455
QY 851 CCATGGCGCGCGCAGACGCTGCTGCGTACGCTGACGCTGATCCCTACGCGCGCTCCACCATCGC 910
Db 456 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 475
QY 911 TCAAGCAGAGACCGCGTGTGCTGCGCGCGAGCACATCATCTGCTGGTGCAGACCCCTTCC 970
Db 476 ProThrThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 495
QY 971 CCGAATGAGCGCGCGCGACCTGATGAATGCGCGCGAGCGCGGCAATCCAAATGATCG 1030
Db 496 LysGluThrAlaProThrThrProLysLysLeuThrProThrProGluLysLeuAla 515
QY 1031 ACGTGTGCGCGAGCTGAGCGCGCGCGCGCGCATCTCATCTGATGAGCAGACCGCGCAG 1090
Db 516 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLeuAlaProThr 533
QY 1091 TGCAGCGCATCTGCGCTGCGCGCGCGCGCATGATCG---GCTCGAGCGCGCTGCGCGCAG 1147
Db 534 ThrProGluLysProThrProThrProThrProGluLysProAlaProThr----- 550
QY 1148 ACGAGCGCGCGCATCGCGCTGCGCGCACTTCCGCGGCGTGGGCGCTGGGCGCATATGCGC 1207
Db 550 ----- 550
QY 1208 GCGAGCTGGGCTGTTCGCCGTGAGACGCGGTATGAGATGACCGCGCTGACGCGCG 1267
Db 551 -----Pro 551
QY 1268 CCGCGTGGCGCTGCGCGCGCGCGCGCGCATGCGAGCGCGGTACTTCGCCGACCTGGTGG 1327
Db 552 LysAlaAlaAla---ProsnThrProLysGluProAlaProThrThrProLysGluPro 570
QY 1328 TGTGAGACCGCGCGACGCGCGCATACG---CCACCTTCGACACACCTACCGCGCGC 1384
Db 571 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 590
QY 1385 CCGCGCGCATCTCCGTGTATGATGACGCGCGCGCGCTGCGAGAGAGCGAGCTCA 1444
Db 591 ProLysGluThrAlaPro---ThrThrLeuLysGluProAlaProThrThrProLysLys 609
QY 1445 CCGCGCAGACGCGCGCGCTGCTGCGACGACGCGCGCGCTGAGCCGCGCGCAGCGCT 1504
Db 610 ProAla-----ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr--- 626
QY 1505 TACATCCGCGGTGAAGGCGCGCGCGCGCGCTCCCAACCTGGAGCAAAACGCT 1564

```

```

Db 627 -----ThrSerAlysProAlaProThrThrProLysGluThrAlaPro 641
QY 1565 ACATGCGCCCTCCCTCCGCTGCAATACGCGCCACCGCATATGCTGGCGCAGAGACTGA 1624
Db 642 ThrThrProLysGluProAlaProThrThrPro----- 652
QY 1625 TGGCGCGCGCGCTCGCGCGCGCGCGCAGAGCGCGGAAATGACCTTCGACAGACCTGCGC 1684
Db 653 -----LysGluProAlaProThrThrProLysLysThrAlaPro---ThrThrLeuLys 669
QY 1685 AGGCGACG---GCATCGCGGTCTCGACCGCTGTCACAG---CCGACCTGGCGC 1732
Db 670 GluProAlaProThrThrProLysLysProAlaProLysGluLeuAlaProThrThrThr 689
QY 1733 AGATCCGCGCTGACCTACGAGACG 1756
Db 690 LysGluProThrSerThrThrSer 697

```

RESULT 9 US-10-124-557-74

Sequence 74, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Alignment Scores: 7.73e-06 Length: 1038
Pred. No.:

Db	480	lys	glu	thr	ala	pro	thr	thr	pro	lys	leu	thr	pro	thr	thr	pro	glu	lys	leu	ala	499	
QY	1031	acg	tgt	tgc	ccg	acg	tgc	agc	ggc	ggc	ggc	ccg	ccg	ccg	ccg	ccg	ccg	ccg	ccg	ccg	1099	
Db	500	pro	thr	thr	pro	-----	glu	ala	pro	ala	pro	thr	thr	pro	glu	leu	ala	ala	pro	thr	517	
QY	1091	tgc	agc	gcg	ac	tct	gac	gtt	gcg	gtt	gcg	ccg	ccg	ccg	ccg	ccg	ccg	ccg	ccg	ccg	1147	
Db	518	thr	pro	glu	glu	pro	thr	pro	thr	pro	thr	pro	glu	glu	pro	ala	ala	pro	thr	thr	534	
QY	1148	acg	agc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	1207	
Db	534	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	534	
QY	1208	ggc	acg	tgc	ggc	ctt	gtt	ctt	ccg	ctg	gag	acg	ggg	gat	gat	gat	acg	ggc	ggc	ggc	1267	
Db	535	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	535	
QY	1268	cgc	gct	tgc	ggc	ctt	gac	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	1327	
Db	536	lys	ala	ala	ala	-----	pro	ala	thr	pro	lys	glu	pro	ala	ala	pro	thr	thr	pro	lys	554	
QY	1328	tgt	tgc	gac	ccg	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	1384	
Db	555	ala	pro	thr	thr	thr	pro	glu	pro	ala	pro	thr	thr	thr	pro	lys	glu	thr	ala	ala	574	
QY	1386	ccg	ggc	ggc	ac	tcc	atc	cat	ccg	tac	gac	gtc	acg	ggc	ggc	ggc	ggc	ggc	ggc	ggc	1444	
Db	575	pro	lys	glu	thr	ala	pro	-----	thr	thr	leu	lys	glu	pro	ala	pro	thr	thr	pro	lys	593	
QY	1445	cgc	ggc	acg	acg	ccg	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	1504	
Db	594	pro	ala	-----	pro	lys	glu	leu	ala	ala	pro	thr	thr	thr	lys	glu	pro	thr	ser	thr	610	
QY	1505	tac	atc	ccg	ggc	ggc	gat	acg	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	1564	
Db	611	-----	-----	-----	-----	-----	thr	ser	ala	lys	pro	ala	pro	thr	thr	pro	lys	glu	thr	ala	625	
QY	1565	acg	atc	ggc	ccg	ctc	ccg	ctc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	1624	
Db	626	thr	thr	thr	pro	lys	glu	pro	ala	pro	thr	thr	thr	thr	thr	thr	thr	thr	thr	thr	636	
QY	1625	tgc	gac	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	1684	
Db	637	-----	lys	glu	pro	ala	pro	thr	thr	pro	lys	glu	thr	ala	pro	-----	thr	thr	leu	lys	653	
QY	1685	agg	ccg	acg	acg	-----	gcg	atc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	ggc	1732	
Db	654	glu	pro	ala	pro	thr	thr	thr	pro	lys	lys	pro	ala	pro	lys	glu	leu	ala	ala	pro	thr	673
QY	1733	agc	atc	ggc	ccg	ctc	gac	ctc	gac	gac	gac	gac	gac	gac	gac	gac	gac	gac	gac	gac	1756	
Db	674	lys	glu	pro	thr	ser	thr	thr	ser	thr	thr	ser	thr	thr	ser							


```

ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseer, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-357-58

Alignment Scores:
Pred. No.: 7,73e-06
Score: 285.50
Percent Similarity: 33.50%
Best Local Similarity: 23.30%
Query Match: 8.65%
DB: 12
Caps: 24

US-10-009-782-1 (1-1758) x US-10-124-557-58 (1-1049)
QY 32 CCATGTCCTCCATCCGATCCAGCCCTTCGACTGTCGCGGGGCGACCTCATCG 91
DB 275 ProthProrThrThrLysSerAlaProThrThrProLysgluProAlaProThr 294
QY 92 ACGGACGCA-----ACACCCGGGGGGGGCGCGCCGACCTGGGCGTGGCGGAC 142
DB 295 ThrlYserAlaProThrThrProLysgluProAlaProThr-----Thr 309
QY 143 GCATCGCCGCGCATCGCATCTGTGCGACGCGCGCGGACACCGCGGGTGCAGTGTGCG 202
DB 310 ThrlYgluProAlaPro-----ThrThrProLysgluProAlaProThr----- 324
QY 203 GCGTGTGTGCGCGCGCGCTTCATCGACTCGACACCGACGACGACACTACCTGCTCA 262
DB 325 -----ThrThrlYsgluProAlaProThrThrThrlYsSer 336
QY 263 GCGCGTCGACATGACGCCAAGATCTGCGAGGGGCTGACACAGGTGTGTCACGGGCAAT 322
DB 337 AlaProThrThr-----ProLysgluProAlaPro----- 346
QY 333 GCGGATACGCGTGGCGCGCTGTGCGACGACACCGCGCGCGCGCGCGCGCGCTGAC 382
DB 347 -----ThrThrProLysgluProAlaProThrThrProLys 358
QY 383 ACGAAGGCGGCTTACCGTTTCAGAGCGCTTGGCGACTACCTGAGACGCGTGGGCGCA 442
DB 359 GluProAlaProThrThrProLysgluProThrProThrThrProLysgluProAlaPro 378

```

```

QY 443 CCGCGGCG-----CCGTACAGCCGCGCTGTATGTGGCCATTCAACCTGGCG 493
DB 379 ThrlYsgluProAlaProThrThrPro----- 388
QY 494 CCGCGGTCATCGCGGACTTCGACGCGCGCGCACGACGAGAAATGCGCGCATGCGG 553
DB 389 -----LysgluProAlaProThrAlaProLysgluProAlaPro 401
QY 554 ACCTGCGCGAGAACGATGCGAGCGCGCGCATTCGACGCGCGCTTAC 613
DB 402 ThrlYsgluProAlaProThrThrProLysgluProAlaProThrThr 419
QY 614 CCGCGCGCGCGCGCGCA-----CCACGAGAGATCATGAGGTGTGCGCGCGTACG 670
DB 420 ThrlYsgluProSerProThrThrProLysgluProAlaProThrThrThrlYsSerAla 439
QY 671 CCGATGGCGGATCATGCGCACCCACAGCGGACGAGGACGACATCGTGGCGCGCG 730
DB 440 ProThrThrThrlYsgluProAlaProThrThrThrlYsSerAlaProThrThrProLys 459
QY 731 TGGAGAAACCTTCGCAATCGCGCGCGAGTGTGCGGTGTGTCGACACA 790
DB 460 GluProSerProThr-----ThrThrlYsgluProAlaProThrThr 473
QY 791 AGGTATGGCGGACCCCAATTTCGCGCGCTGCGGAGACGCTGCGGTATCGAGCGCG 850
DB 474 ProLysgluProAlaPro-----ThrThrPro 482
QY 851 CCATGGCGCGCGACGCTGCTGAGCGGTATCCCTACGTGCGCGCGTCCACATGC 910
DB 483 LyslYsProAlaProThrThrProLysgluProAlaProThrThrProLysgluProAla 502
QY 911 TCAAGCAGAACCGCGTCTGTCGCGCGACCATCATCATCTGATGACGACCTTC 970
DB 503 ProThrThrThrlYsLysProAlaProThrAlaProLysgluProAlaProThrPro 522
QY 971 CCGAATGACGCGCGCGGACCTGATGAGTCCGCGCGGACGCGGCAATTCAGATGAC 1030
DB 523 LysgluThrlAlaProThrThrProLyslYsleuThrProThrThrProgluYsleuAla 542
QY 1031 ACGTGTGCGCGAGTGCAGCGCGCGCGGACATCATCTATGATGACGAGAACCGGACG 1090
DB 543 ProThrThrPro-----GluYsProAlaProThrThrProgluYsleuAlaProThr 560
QY 1091 TGGAGGCAATCTGCGGTGCGCGCGACCATGATG-----GCTCCGACGCGCGCGGACG 1147
DB 561 ThrlYsgluProThrProThrThrProgluYsleuAlaProThr----- 577
QY 1148 ACGAGCGCGCGCATCGCGCGCTGTGGGACCTCCCGCGGTGTGGGCGACTATGCGC 1207
DB 577 ----- 577
QY 1208 GCGACCTGGCGCTTCCCGCTGTGAGACGCGGATGAGATGACCGGCTGACCGCG 1267
DB 578 -----Pro 578
QY 1268 GCGGCTTGACCTGGCGCGCGCGCGGACGCTGACGCGCGGTACTGCGCGACCTGTGG 1327
DB 579 LysAlaAlaAla-----ProAlaThrProLysgluProAlaProThrThrProLysgluPro 597
QY 1328 TGTTCGACCGCGCGCGCGCGATACCG-----CCACTTCGAAACCTTCAGAGCGCG 1384
DB 598 AlaProThrThrProLysgluProAlaProThrThrProLysgluThrlAlaProThrThr 617
QY 1385 CCGCGCGCATCATCTCGGTATGCTGACGACGCGCGCGGTGTGCAAGACGACGCTTCA 1444
DB 618 ProLysgluThrlAlaPro-----ThrThrlYsgluProAlaProThrThrProLys 1444
QY 1445 CCGCGGAGATGCGCGCGCGGTGCTGCGACGACGCGCGCGCGCGCGCGCGCTTAC 1504
DB 637 ProAla-----ProLysgluYsleuAlaProThrThrThrlYsgluProThrSerThr 653
QY 1505 TACATCGCGGTGAACGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCT 1564

```


[illegible]

```

QY 443 CGCGGCGG-----CCGTCAACGCCGCTGTATGATGGCCATTCAACGTCGCG 493
Db 336 ThThrLysGluProAlaProThrThrPro----- 345
QY 494 CGCGGTCATGCGGCACTGACGCGCGCCAGAGAGAAATCGCGGCATCGCGG 553
Db 346 -----LysGluProAlaProThrThrThrProLysProAlaPro 358
QY 554 ACCTGCGCGAGAGAACCATGGCCAGCGCGCCATTCGACCGCGCTTCAC 613
Db 359 ThThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 376
QY 614 CGCCGCGCGCGCGCA-----CCACCGAAGATCATGAGTGTGCGCGCGCTGACG 670
Db 377 ThThrLysGluProSerProThrThrProLysGluProAlaProThrThrThrLysSerAla 396
QY 671 CGCATGCGCGCATTCACCCACCATGACGCGAGAGCGGACACATCGCGCGG 730
Db 397 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 416
QY 731 TGGAGGAACCTTCGCGCATCGCGCGCGACGTCGCGGTGATCGACACACA 790
Db 417 GluProSerProThr-----ThThrLysGluProAlaProThrThr 430
QY 791 AGGTATGCGCGCACCCCAATTTCGCGCGCTGCGGAGACGCTGCTGATCGAGCGG 850
Db 431 ProLysGluProAlaPro-----ThThrPro 439
QY 851 CCATGCGCGCGCACCATCTGCTGAGACGCGTATCCCTACGTCGCGCGCTCAACATGC 910
Db 440 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 459
QY 911 TCAGAGGAGCGCGCTGCTGCTGCGCGCGACGACCATCATCAGCTGTCAGAGCGCTTCC 970
Db 460 ProThrThrThrLysLysProAlaProThrThrThrProLysGluProAlaProThrThrPro 479
QY 971 CCGACATGACGCGCGCGCATGATGATGAGTCGCGGCGCGGCAATCCAGATACG 1030
Db 480 LysGluThrAlaProThrThrProLysLysLysLysLysLysLysLysLysLysLysLysLys 499
QY 1031 ACCTGTCGCGCGACGTCGACGCGCGCGCGCATCTATCATGATGAGAGACGCGAGC 1090
Db 500 ProThrThrPro-----GluLysProAlaProThrThrProLysGluLysLysLysLysLys 517
QY 1091 TGCAGGAGCATTCGCGGTGCGCGCGACCATGATTCG-----GCTCCGACGCGCTCGCGCAGC 1147
Db 518 ThProGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 534
QY 1148 ACGAGCGCGCGCATCCGCGCTGTGGGACACTTCCCGGCGGTGTCGGGCGCACTATGCGC 1207
Db 534 ----- 534
QY 1208 GCGACTGCGGCTGTCCCGCTGAGAGCGCGGTATGAAATGACCGCGCTGACGCGCG 1267
Db 535 -----Pro 535
QY 1268 CCGCTTCGCGCTGCGCGCGCGCGCGCGAGCTGAGCGCGGTACTTCCCGACCTGCTG 1327
Db 536 LysAlaAlaAlaAla-----ProsmThrProLysGluProAlaProThrThrProLysGluPro 554
QY 1328 TGTTCGACCGCGCGAGGTCGCGCATACCG-----CCACCTTCGACACCTACCGAGCGG 1384
Db 555 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 574
QY 1385 CCGCGCGCATTCCTCGCTGACGTCACAGCGCGCGCGCTGCGCAAGAGCGGCTTCA 1444
Db 575 ProLysGlyThrAlaPro-----ThThrLysLysGluProAlaProThrThrProLysLys 593
QY 1445 CCGGCGCAGATCGCGCGCGCTGCGCGAGACGCGCGCTGAGCGCGCGCGCGCGCGCT 1504
Db 594 ProAla-----ProLysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 610

```

```

QY 1505 TACATCCGCGGTGAGAGGCGCGGTGCGCGCGCTCCCAACCTGAGACCAACCGCT 1564
Db 611 -----ThSerAspLysProAlaProThrThrProLysGlyThrAlaPro 625
QY 1565 ACATGCGCGCTCCCTCCCTGCGCATACGCGCGCGCCACCGCATATGTCGGGCAAGAAATGA 1624
Db 626 ThThrProLysGluProAlaProThrThrPro----- 636
QY 1625 TGGGCGCGCGCTGCGCGCGCGCGCAAGCGCGCGGAAATGACCTGCAAGACCTGTCG 1684
Db 637 -----LysGluProAlaProThrThrProLysGlyThrAlaPro-----ThThrLysLys 653
QY 1685 AGCGACGCG-----GCATCGCGCTGTCGACCCCTGTCAGG-----CCGAGCTGGCGC 1732
Db 654 GluProAlaProThrThrProLysLysProAlaProLysGluLysLysLysLysLysLysLys 673
QY 1733 AGATGCGCGCTGAGTCAGAGAGC 1756
Db 674 LysGlyProThrSerThrThrSer 681

```

RESULT 13

US-10-124-557-42

Sequence 42, Application US/10124557

Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENGTH: 1311 amino acids

TYPE: amino acid

MOLECULE TYPE: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 42:

US-10-124-557-42

Alignment Scores:

Pred. No.: 7.56e-06 Length: 1311
 Score: 285.50 Matches: 137
 Percent Similarity: 33.50% Conservative: 60
 Best Local Similarity: 23.30% Mismatches: 240
 Query Match: 8.65% Indels: 151
 DB: 12 Gaps: 24

US-10-009-782-1 (1-1758) x US-10-124-557-42 (1-1311)

QY 32 CCATGCCCAATCCGATCCAGCCCTTGCACCTGCTGCGGGCGGCACCTCATCG 91
 DB 273 ProthPrProthPrThrLysSerAlaProthPrThrProLysGluProAlaProthPrThr 292
 QY 92 AGGCGACA-----ACACCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 142
 DB 293 ThrLysSerAlaProthPrThrProLysGluProAlaProthPrThr-----Thr 307
 QY 143 GCATGCGCGCCATCGGCATCTGTGAGACCGCGCGGCACACCGGGGTGACGTCGCG 202
 DB 308 ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProthPr----- 322
 QY 203 GCGTGTGTGTCGCGCGGCTTCATCGACTGCGACACCGACGAGACACCACTACTGCTCA 262
 DB 323 -----ThrThrLysGluProAlaProthPrThrProLysSer 334
 QY 263 GCGCGCGACATGACGCCCAAGATCTGCGAGGGGCTGACCGAGGTGTGTCAGGGCAATT 322
 DB 335 AlaProthPr-----ProLysGluProAlaPro----- 344
 QY 323 GCGGCATGACCTGGCGCGCGTGGGACACCGCCACCGCCCGCCCTGACCTGCTGG 382
 DB 345 -----ThrThrProLysGluProAlaProthPrThrProLys 356
 QY 383 ACGAAGGCGCTTACCGCTTTCAGCGCTTCGCGGACTACCTGAGCGCTTGGGGCCA 442
 DB 357 GluProAlaProthPrThrProLysGluProthPrThrProLysGluProAlaPro 376
 QY 443 CGCGCGCG-----CGTCAACCGCGCTGTATGTTGGGCATTCACGCTGCGCG 493
 DB 377 ThrThrLysGluProAlaProthPrThrPro----- 386
 QY 494 CGCGGTCATGCGGACTTCAGCGCGCGCCACCGACGAGAAATCGCGCCATGCGGG 553
 DB 387 -----LysGluProAlaProthPrAlaProLysLysProAlaPro 399
 QY 554 ACCTGGCGGAGAGACGATGCGCGCGCCATCGGACATTCGACGCGCGCTTCTACG 613
 DB 400 ThrThrProLysGluProAlaProthPrThrProLys-----GluProAlaProthPrThr 417
 QY 614 CGCGCGCGCGCGCA-----CGACGAGAGATCATCGAGGTGTCGCGCGCTGAGCG 670
 DB 418 ThrLysGluProSerProthPrThrProLysGluProAlaProthPrThrLysSerAla 437
 QY 671 CGCATGCGGATCTACGACCCACCATCGCGAGAGCGAGCGACATCGGCGCGCG 730
 DB 438 ProthPrThrThrLysGluProAlaProthPrThrLysSerAlaProthPrThrProLys 457
 QY 731 TGGAGAAACCTTCGCGATCGCGCGCGAGCTGAGTGGTGTGATCTGACACACA 790
 DB 458 GluProSerProthPr-----ThrThrLysGluProAlaProthPrThr 471
 QY 791 AGGTATGCGGACGCAATTTCGCGCGCTGCGCGGACAGAGCTGCGCTGATCGAGCGG 850
 DB 472 ProLysGluProAlaPro-----ThrThrPro 480
 QY 851 CCATGCGCGCGGACGATCTGCGTGAAGCGTATCCCTAGTGCGCGGCTCCACACATG 910
 DB 481 LysLysProAlaProthPrThrProLysGluProAlaProthPrThrProLysGluProAla 500
 QY 911 TCAACAGAGAGCGCTGCTGCGCGGACGACCATCATCACTGTGTGACAGCCCTTCC 970
 DB 501 ProthPrThrThrLysLysProAlaProthPrThrAlaProLysGluProAlaProthPrThrPro 520

QY 971 CCGAATGAGCGGGCGGACCTGATGAGTGGCGGGCCGAGCGGCAATTCAGTACG 1030
 DB 521 LysGluThrAlaProthPrThrProLysLysLeuThrProthPrThrProLysLeuAla 540
 QY 1031 ACGTGTGCGCGAGCTGACGCGCGCGCGCCATCTACTCATGTATGACGAAACCGAG 1090
 DB 541 ProthPrPro-----GluLysProAlaProthPrThrProLysGluLeuAlaProthPr 558
 QY 1091 TGCACGATCTGTGCGCTTGGCGCGGACCATGATCG-----GCTCGAGCGGCTGCGCAG 1147
 DB 559 ThrProGluGluProthPrProthPrThrProLysGluProAlaProthPrThr----- 575
 QY 1148 ACGAGCGCGCCATCCGCGCTGTGGGGCACCTTCCGCGGGGTCTGGGGCACTATGCGC 1207
 DB 575 ----- 575
 QY 1208 GCGACCTGGCGCTTCCCGCTGAGACGCGGTATGAGATGACCGGCTGACCGCG 1267
 DB 576 -----Pro 576
 QY 1268 GCGCTTGGCGCTTGGCGCGCGGCGGACGTCGACGCGCGGATCTTGGCGACCTGTGG 1327
 DB 577 LysAlaAlaAla-----ProAsnThrProLysGluProAlaProthPrThrProLysGluPro 595
 QY 1328 TGTTCGACCGCGCGCGCGGATACCG-----CCACCTTCGAAACACCTTACGAGCGCG 1384
 DB 596 AlaProthPrThrProLysGluProAlaProthPrThrProLysGluThrAlaProthPrThr 615
 QY 1385 CCGCGCGCATTCATCCGCTTACGTCAACGCGCGCGCGGTGTCGACAGACGAGCGTTCA 1444
 DB 616 ProLysGluThrAlaPro-----ThrThrLeuLysGluProAlaProthPrThrProLys 634
 QY 1445 CGCGGACGATGCGCGCGCGGCTGTCGACGACGAGCGCGCGCTGACGCGCGCGCCCT 1504
 DB 635 ProAla-----ProLysGluLeuAlaProthPrThrThrLysGluProthPrThr 651
 QY 1505 TCAATCGCGGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1564
 DB 652 -----ThrSerAspLysProAlaProthPrThrProLysGluThrAlaPro 666
 QY 1565 ACATGCGCGCTCCCTCCCTCGCGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624
 DB 667 ThrThrProLysGluProAlaProthPrThrPro----- 677
 QY 1625 TGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1684
 DB 678 -----LysGluProAlaProthPrThrProLysGluThrAlaPro-----ThrThrLeuLys 694
 QY 1685 AGCGCAGG-----GCATCGCGGCTCTGACCTGTGCAAG-----CCGAGCTGGCGC 1732
 DB 695 GluProAlaProthPrThrProLysLysProAlaProLysGluLeuAlaProthPrThr 714
 QY 1733 AGATGCGCGCTGAGTACGAGAGAGC 1756
 DB 715 LysGluProthPrSerThrThrSer 722

RESULT 14

US-10-124-557-142
 Sequence 142, Application US/10124557
 Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Gesser, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESS: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge

```

STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 1313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142
Alignment Scores:
Pred. No.: 7.56e-06 Length: 1313
Score: 285.50 Matches: 137
Percent Similarity: 33.50% Conservative: 60
Best Local Similarity: 23.30% Mismatches: 240
Query Match: 8.65% Indels: 151
Gaps: 24
US-10-009-782-1 (1-1758) x US-10-124-557-142 (1-1313)
QY 32 CGATGCCCAATCCGATTCGACCCCTGCTGCGGGCGGCACCCCTCATCG 91
DB 275 ProthProthThrThrLysSerAlaProthThrThrProlysgluProAlaProthThr 294
QY 92 ACGGAGCA-----ACACCCGGGGGGGGCGCGACCTGGGGCGTGGCGGAC 142
DB 295 ThrlysserAlaProthThrProlysgluProAlaProth-----Thr 309
QY 143 GCATGCCCGCATCGCGATCTGCGACGCGCGCGACACACCGCGGTGACGTGCGG 202
DB 310 ThrlysgluProAlaPro-----ThrThrProlysgluProAlaProth----- 324
QY 203 GCGTGTGTGTCGGCGCGCTCATCGATCGACACCAACGACGACAACTACCTGCTCA 262
DB 325 -----ThrThrlysgluProAlaProthThrThrThrlysser 336
QY 263 GCGGTGCGACATGACGCCAGATCTGCGAGGGGTGACACGGGTGTCACGGGCAATT 322
DB 337 AlaProthThr-----ProlysgluProAlaPro----- 346
QY 323 GCGGATACGCTGGCGCGCTGCGCACGCCACCCCGCGCTGACCTGCTGG 382
DB 347 -----ThrThrProlysgluProAlaProthThrProlysglu 358
QY 383 ACGAAGGCGGCTTACGCTTCGAGCGCTTGCGGACTGACGCGGTGCGGGCCA 442
DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

```

```

DB 359 gluProAlaProthThrThrProlysgluProthProthThrProlysgluProAlaPro 378
QY 443 CGCCGCGCG-----CGGTCAACGGCGGCTGATGTGGGCATTCAACGCTGCGCG 493
DB 379 ThrThrlysgluProAlaProthThrPro----- 388
QY 494 CCGGCGTCATCCGGACTGTGACGCGCGCGCGACCGACGAGAAATGCGGCGCATGGCGG 553
DB 389 -----LysgluProAlaProthThrAlaProlysgluProAlaPro 401
QY 554 ACCTGCGCGAGAAACCATGCGCACGGCGCCATGCGCATTCGACCGGCGCTTCAC 613
DB 402 ThrThrProlysgluProAlaProthThrProlysgluProAlaProthThr 419
QY 614 CGCCGCGCGCGCGCGCA-----CAACGAGAGATCAGTCAGAGTGTGCGCGCGTGAAGC 670
DB 420 ThrlysgluProserProthThrThrProlysgluProAlaProthThrThrLysSerAla 439
QY 671 CGCATGCGCGCATCTACGCCACCCACATGCGCGAGAGAGGCGACATGCGCGCGCG 730
DB 440 ProthThrThrlysgluProAlaProthThrThrlysserAlaProthThrProlysglu 459
QY 731 TGGAGGAACCTTCCGATGCGCGCGCGACGTCGCGGTGATGTCGACACCA 790
DB 460 GluProserProth-----ThrThrlysgluProAlaProthThr 473
QY 791 AGTCATGCGCGACCCCAATTCGCGCGCTGCGCGAGAGCGTGCCTCATGAGCGCG 850
DB 474 ProlysgluProAlaPro-----ThrThrPro 482
QY 851 CCATGCGCGCGCGACGCTGCTGCGACGCGATCCCTACGTGCGCGCGCTGACACATGC 910
DB 483 LyslysgluProAlaProthThrProlysgluProAlaProthThrProlysgluProAla 502
QY 911 TCAAGCAGACCGCGTCTGCTGCGCGAGCAGACATCACCCTGCTGACACCTTCC 970
DB 503 ProthThrThrlysgluProAlaProthThrAlaProthThrProlysgluProAlaProthThrPro 522
QY 971 CCGAATGACGCGCGCGCGCGCTGATGATGCGCGCGCGCGCGCAATTCACAGTAGC 1030
DB 523 LysgluThrAlaProthThrProlysgluProlysgluThrProthThrProgluLysleuAla 542
QY 1031 ACGTGTGCGCGCGACGTGACCGCGCGCGCGCGCATCTACTCATGATGAGACGCGAGC 1090
DB 543 ProthThrPro-----gluLysProAlaProthThrProgluLysleuAlaProthThr 560
QY 1091 TGCAGGCATCTGCGCGCGCGCGCGACATGATCG-----GTCGCGAGCGCGCGCGACG 1147
DB 561 ThrProgluThruProthThrProthThrProgluThruProgluThruProthThr----- 577
QY 1148 ACGAGCGCGCGCATCGCGCGCTGTGGGACATTCGCGCGGTGTGGGGACTATGCGC 1207
DB 577 ----- 577
QY 1208 GCGACTGCGCGCTGCTCCGCTGCGAGACGCGGTATGAAATGACCGGCTGACCGCG 1267
DB 578 -----Pro 578
QY 1268 CGCGCTTGGCGCTGCGCGCGCGCGCGACCTGACAGCGCGGATTCGCGCGACCTGTGG 1327
DB 579 LysAlaAlaAla-----ProsnThrProlysgluProAlaProthThrProlysgluPro 597
QY 1328 TGTTCGACCGCGCGACGATGCGCATACG-----CCACCTTCAGACACCTTACGAGCGG 1384
DB 598 AlaProthThrProlysgluProAlaProthThrProlysgluThrAlaProthThr 617
QY 1385 GCGCGCGCATTCCTGCTACGTCAACGCGCGCGCGCTGCGCAAGACGCGTTCA 1444
DB 618 ProlysgluThrAlaPro-----ThrThrleuLysgluProAlaProthThrProlysglu 636
QY 1445 CCGGCGAGCATGCGCGCGCGGTGTCGACGACGAGCGCGCGCGCGCGCGCGCGCGCT 1504
DB 637 ProAla-----ProlysgluLysleuAlaProthThrThrLysgluProthThrSerThr--- 653

```

MOLECULE TYPE: protein.
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
IS-10-134-557-50

[illegible]

Db 504 ProThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 523
QY 971 CCGAAGTGGAGCGCGCGCGAGCTGGATGAAGTCCGCGCGCGCGCGCAATCCAGTAGC 1030
Db 524 LysGluThrAlaProThrThrProLysLysLeuThrProThrThrProGluLysLeuAla 543
QY 1031 ACGTGTGCGCGCGAGTGCAGCGCGCGCGCGCATCTCATGATGAGAGAACCGCAGC 1090
Db 544 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLeuAlaProThr 561
QY 1091 TGCAGCGCATCTCGGCGTCCGCGCGCGCATGATCG--GCTCCGAGCGCGCTGCGCAGC 1147
Db 562 ThrProGluGluProThrProThrProThrProGluGluProAlaProThrThr----- 578
QY 1148 ACGAGCGCGCGCATCGCGCGCTGCGGACCTTCCCGCGGCTGCGGCGCTATGCGC 1207
Db 578 ----- 578
QY 1208 GCGACCTGGGCGCTGTCCCGCTGAGACGCGCGGTATGAAGATGACCGCGCTGACCGCG 1267
Db 579 -----Pro 579
QY 1268 GCGCGCTTGGCGCTGGCGCGCGCGCGAGCTGAGCGCGGCTGACTTCCGCGACCTGCTGG 1327
Db 580 LysAlaAlaAla---ProAsnThrProLysGluProAlaProThrThrProLysGluPro 598
QY 1328 TGTTCGACCGCGCGCGCGCGCGCGATACG---CCACCTTCGACACCTACCGAGCGCG 1384
Db 599 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 618
QY 1385 CCGCGCGCATTCATTCGCTGATGATGACGCGCGCGCGCGCGCGCGCAAGACGCGCTCA 1444
Db 619 ProLysGluThrAlaPro---ThrThrLeuLysGluProAlaProThrThrProLysLys 637
QY 1445 CCGCGCGAGCATCCGCGCGCGCTGCTGCGACGACGCGCGCGCGCGCGCGCGCGCT 1504
Db 638 ProAla-----ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr--- 654
QY 1505 TACATCCGCGCGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1564
Db 655 -----ThrSerAspLysProAlaProThrThrProLysGluThrAlaPro 669
QY 1565 ACATGCGCCCTCCCTCCGCTCGCATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624
Db 670 ThrThrProLysGluProAlaProThrThrPro----- 680
QY 1625 TGGCGCGCGCGCTGCT 1684
Db 681 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLeuLys 697
QY 1685 AGCCAGCG---GCATCGCGCGCTGACCGCTTCCAGG-----CCGAGCTGGCGC 1732
Db 698 GluProAlaProThrThrProLysLysProAlaProLysGluLeuAlaProThrThrThr 717
QY 1733 AGATCGCCCTGAGTAGAGAGC 1736
Db 718 LysGluProThrSerThrThrSer 725

Search completed: May 11, 2003, 12:30:20
Job time : 96.5 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: May 11, 2003, 12:10:05 ; Search time 36.5 Seconds

(without alignments)
2834.274 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 3299
Sequence: 1 gaattccactgacgcgca.....ccctgagctacgagaagcct 1758

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 26574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPRO_sepool/US10009782/runat_07052003_122518_23179/app_query.fasta_1.1927
-DB=Issued_Patents_AA -OEXT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=Dits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10009782 -ECGN 1.1.21 -runat 07052003_122518_23179 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEDQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEROUT=120
-WARN_TIMEROUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	398.5	12.1	1213	4	US-09-413-814-79
2	344	10.4	882	4	US-09-413-814-78
3	320.5	9.9	882	4	US-09-413-814-78
4	310	9.4	558	4	US-09-199-637A-277
5	309	9.5	1213	4	US-09-413-814-79
6	287.5	8.7	960	4	US-09-219-849-6
7	285.5	8.7	941	4	US-07-757-022B-14
8	285.5	8.7	1022	4	US-07-757-022B-84
9	285.5	8.7	1038	4	US-07-757-022B-84
10	285.5	8.7	1049	4	US-07-757-022B-58
11	285.5	8.7	1140	4	US-07-757-022B-104
12	285.5	8.7	1270	4	US-07-757-022B-44

13	285.5	8.7	1311	4	US-07-757-022B-42	Sequence 42, Appl
14	285.5	8.7	1313	4	US-07-757-022B-142	Sequence 142, Appl
15	285.5	8.7	1314	4	US-07-757-022B-50	Sequence 50, Appl
16	285.5	8.7	1320	4	US-07-757-022B-46	Sequence 46, Appl
17	285.5	8.7	1330	4	US-07-757-022B-60	Sequence 60, Appl
18	285.5	8.7	1354	4	US-07-757-022B-48	Sequence 48, Appl
19	285.5	8.7	1361	4	US-07-757-022B-40	Sequence 40, Appl
20	285.5	8.7	1363	4	US-07-757-022B-52	Sequence 52, Appl
21	285.5	8.7	1404	4	US-07-757-022B-2	Sequence 2, Appl
22	285.5	8.7	1404	4	US-07-757-022B-62	Sequence 62, Appl
23	283	8.6	1065	1	US-08-642-255-72	Sequence 72, Appl
24	277	8.4	960	4	US-09-219-849-5	Sequence 5, Appl
25	276.5	8.4	1057	3	US-08-831-820-1	Sequence 1, Appl
26	275	8.3	1064	1	US-08-642-255-62	Sequence 62, Appl
27	270	8.2	1461	4	US-09-585-887-9	Sequence 9, Appl
28	270	8.2	1461	4	US-09-289-578-9	Sequence 9, Appl
29	267.5	8.2	1958	1	US-07-945-283-2	Sequence 2, Appl
30	267	8.1	720	4	US-09-219-849-4	Sequence 4, Appl
31	267	8.1	777	1	US-08-642-255-53	Sequence 53, Appl
32	267	8.1	1341	3	US-08-663-825-18	Sequence 18, Appl
33	267	8.1	1341	4	US-09-500-811-18	Sequence 18, Appl
34	267	8.1	1341	4	US-09-570-573-18	Sequence 18, Appl
35	267	8.1	1341	4	US-09-548-608-18	Sequence 18, Appl
36	266	8.1	552	4	US-09-219-849-7	Sequence 7, Appl
37	263	8.0	822	4	US-09-219-849-49	Sequence 49, Appl
38	262.5	8.0	633	1	US-08-642-255-73	Sequence 73, Appl
39	262	7.9	1958	1	US-07-945-283-2	Sequence 2, Appl
40	257.5	7.9	355	4	US-08-483-533-41	Sequence 41, Appl
41	257.5	7.9	355	4	US-09-283-471A-41	Sequence 41, Appl
42	257.5	7.9	355	5	PCT-US91-06532-3	Sequence 3, Appl
43	254.5	7.7	395	4	US-09-370-838-187	Sequence 187, Appl
44	252	7.6	1442	2	US-08-316-650-12	Sequence 12, Appl
45	252	7.6	1442	5	PCT-US95-02251-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-413-814-79
Sequence 79, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloeker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoffe, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
EARLIER FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 79
LENGTH: 1213
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-79
Alignment Scores:
Pred. No.: 1,416-19
Score: 398.50
Percent Similarity: 36.428
Best Local Similarity: 32.258
Length: 1213
Matches: 209
Conservative: 27
Mismatch: 230

```

Qy 941 GCACCATCATTACCGTGTGCAAGCCCTTCCCGCACTGAGCGGGCGCACCGATGAAG 1000
Db 328 -----ProArgArgaspProHisArg----- 334

```

[illegible]

RESULT 2
 US-09-413-814-78
 Sequence 78, Application US/09413814
 Patent No. 6225064
 GENERAL INFORMATION:
 APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
 APPLICANT: Bristol-Myers Squibb, Co.
 APPLICANT: Beyer, Stefan
 APPLICANT: Bloeker, Helmut

Db 691 rglYAlaAlaGluProArgArgArgArgArgAlaAlaProProArgProArgInglYAlaLeuA 711
 QY 1548 -----CCTGACGCAACCCCTACATGATGCCCCCTCCCTCGCTGCAATAGCGCCACCC 1602
 Db 711 rglalglYAspLeuProLeuHISglYProAlaArgAlaGlnAlaArgPro----- 729
 QY 1603 GATATCGTGGCAAGAAAGTATGGGGCGCCCTGGCGCCGCGCAAGCGCAAGCGCGAA 1662
 Db 730 -----AlaProAlaHISArgAlaArgGlyProArgA 740
 QY 1663 ATGACCCCTGCAGACCTGTGCGAGCGCGATGCGGCTCGACCTGTCCAGGCC 1722
 Db 740 rG-----GlyHISArgArgArgGluProGluProArgArgAspA 753
 QY 1723 GAGCT-----GGCCAGAT-----CGCC 1740
 Db 753 rglAlaAspArgArgLeuArgGlnProAlaArgAlaProAspArgLeuArgGlyArgP 773
 QY 1741 CTGAGCTACGAGAAAGCT 1757
 Db 773 roAspValArgArgAla 778

RESULT 3

US-09-413-814-78
 ; Sequence 78, Application US/09413814
 ; Patent No. 6225064
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
 ; APPLICANT: Bristol-Myers Squibb, Co.
 ; APPLICANT: Beyer, Stefan
 ; APPLICANT: Bloeker, Helmut
 ; APPLICANT: Cino, Paul M
 ; APPLICANT: Dougherty, Brian A
 ; APPLICANT: Goldberg, Steven L
 ; APPLICANT: Hofle, Gerhard
 ; APPLICANT: Mueller, Joachim
 ; APPLICANT: Reichenbach, Hans
 ; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
 ; FILE REFERENCE: PCT/US 99/23535
 ; CURRENT APPLICATION NUMBER: US/09/413,814
 ; EARLIER FILING DATE: 1999-10-07
 ; EARLIER APPLICATION NUMBER: DE 198 46 493.2
 ; NUMBER OF SEQ ID NOS: 107
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 78
 ; LENGTH: 882
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 US-09-413-814-78

Alignment Scores:

Pred. No.: 2,64e-14 Length: 882
 Score: 320.50 Matches: 204
 Percent Similarity: 34.43% Conservative: 16
 Best Local Similarity: 31.92% Mismatches: 260
 Query Match: 9.88% Indels: 159
 DB: 4 Gaps: 32

US-10-009-782-1 (1-1758) x US-09-413-814-78 (1-882)

QY 1735 TCTGGCCAGCTCGGCTTGACAGGCTGACAGCCGATCCGCTGCGCTGCGCAAGCT 1676
 Db 301 AlaelYHISAlaValaAspTPrThr---AlaArgProArgGlyArgProProAlaArgSer 319
 QY 1675 CTTCGA-----GGGTCAATTTCCGGGCTTGCGCTGCGCGCGAGCGCG 1631
 Db 320 SerAlaCysArgArgThArgSerSerGlyArgGlyThArgSerArgArgAlaAla 339
 QY 1630 CGCCATCATTCTCTTCCCAAGATATCGGTGGGCGCGTATTCGAGCGGAGGAGGG 1571

Db 340 ProProPro---ProSerArgArgArgProThArgProSerArgProArgArgThPro 358
 QY 1570 CCATGTAGCCGTTTGCTCCAGAGGTGGAGGGCGCGACGCGCCCGCTTCACGC--- 1514
 Db 359 SerArgArgArgArgSerProAlaArgThProGlyProArgSerProArgArgPro 378
 QY 1513 ---CGATTGTAGGGCTGGCGGCTCAGCGCGCTGCGTGCAGACCGCGCGCG 1457
 Db 379 SerProSerAlaArgSerProAlaAlaArgArgProProArgArgProArgArgArg 398
 QY 1456 CATGCTGGCGCGTGAACGCGCTGCTGCGCAGACGCGCGCGCTGACGTACAGGAAT 1397
 Db 399 ArgProArgArg---LeuLeuArgProAlaArgArgLeuAlaArgArgAsp 414
 QY 1396 -----GGATCCGCGCGCGCT----- 1379
 Db 415 ProGlyAspValProAspProArgAlaAlaArgArgArgAlaProAlaAlaLeu 434
 QY 1378 -----CGGTAGGCTGTGCAAGGTGGCGGTATCGGCCACCGTGGCGGT 1334
 Db 435 ProGlyAlaAspGlyArgGlyAlaArgGlyProGlyArgArgArgAlaProArgArg 454
 QY 1333 -----CGAACACACACAGTGC-----CGAAGTACCGCGCGCTCAGCT 1295
 Db 455 AlaAlaAlaArgValProProAspArgAlaAspProProAlaArgProAlaAlaVal 474
 QY 1294 GCCCGCGCGCGCGCAGC-----CGAACCGCGCGCGCTCAGCGCTCATCTTC 1244
 Db 475 ValArgAlaAlaProAlaAlaValaValaArgArgSerAlaArgAlaGlnProGlyAlaGln 494
 QY 1243 ATACCGCGCTCTCCAGCGGGAACAGCCGAGTGGCGCGCGCATAGCCCA----- 1193
 Db 495 HisProValArgGlyAlaArgAspGlyProAlaArgArgArgArgAlaProProGlnPro 514
 QY 1192 -----GCACCGCGGGAAGTGCCTCCACAGCGCGGATGCGCGCTGT 1148
 Db 515 ValArgAspArgAlaAlaProProArgGlyAlaAlaAspArgValGlnArgGlnArgArg 534
 QY 1147 CGTGGCGACCGCGTGGAGCCGATCATGCTCGCGCGCAACGCGAGTGCCTGCACGT 1088
 Db 535 AlaValProGlyArgArgAlaArgGly-----ProGly---AlaValPro 548
 QY 1087 CGGCTCGCTCATCATGATGATGATGCGCGCGCGCTGACCTGCGGCGACAGCTGT 1028
 Db 549 AspValArgPro-----ArgAlaProArgGlyArgGlyProArgArg 562
 QY 1027 ACTTGATTTGCGCGCTGCGCGCTCATCTCATCGTCCGCTCAGTTCGAGA 968
 Db 563 GlyGlyLeu---GlyAlaArgProArg-----GlyGluProArgAlaValaArg--- 577
 QY 967 AGGCTTGACACAGGTATGATGTGCGCGCGCGCGCGCGCGCTCTCTTGAGCA 908
 Db 578 -----ProGluProArgAlaAlaProProArgAlaGly 588
 QY 907 TGTGTGACCGCGCCAGTATGAGGTATGCGTCCAGCGAGACGTCTGGCGCGCGAGCGG 848
 Db 589 HisProGluAlaArgArgArgAlaArgHisArgAlaArgGly-----AlaProArgArg 606
 QY 847 CCTGATAGAGGCA-----GGTCTCGCGCGAGCGCGCGCAAT 809
 Db 607 ValArgArgLeuValGlyArgArgLeuArgArgArgArgAlaArgAlaLeuArgArgLeu 626
 QY 808 TGGCTGCGCCATACCTGTGTGCGAGATACACCGCGAGCTCCAGCTGCGCGCGCA 749
 Db 627 ArgAlaGly-----ProAlaArgProAlaAlaGlyAla 637
 QY 748 TCGGGAAGTTCTCTCAGCGCG-----CCACGATGTGTGCTCTGCGCGCGCATGT 695
 Db 638 ProGlyAlaValaValaArgArgArgArgSerProAlaGlyAlaAlaValaArgArgGlyPro 657
 QY 694 GGTGTGGGTAGTACCGCGCATGCGCGCTCAGCGCGCGCACACCTGATGATCTCTTGG 635
 Db 658 GlyGlyArgAlaProValleuAspArgGluAlaLeuGly-----Arg 671

APPLICANT: Brandt, Petra
 APPLICANT: Cino, Paul M
 APPLICANT: Dougherty, Brian A
 APPLICANT: Goldberg, Steven L
 APPLICANT: Hotte, Gerhard
 APPLICANT: Mueller, Joachim
 APPLICANT: Reichenbach, Hans
 TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
 FILE REFERENCE: PCT/US 99/23533
 CURRENT APPLICATION NUMBER: US/09/413,814
 EARLIER FILING DATE: 1999-10-07
 EARLIER APPLICATION NUMBER: DE 198 46 493.2
 NUMBER OF SEQ ID NOS: 107
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO: 78
 LENGTH: 882
 TYPE: PRT
 ORGANISM: Sorangium cellulosum
 US-09-413-814-78

Alignment Scores:
 Pred. No.: 6,65e-16 Length: 882
 Score: 344.00 Matches: 196
 Percent Similarity: 34.68% Conservative: 36
 Best Local Similarity: 29.30% Mismatches: 214
 Query Match: 10.43% Indels: 23
 Gaps: 33

US-10-009-782-1 (1-1758) x US-09-413-814-78 (1-882)

49 TCCAGACCCCTTCGACCTGCTCGGGGGGACCCCTTCAGCGCAGACACCCCG 108
 237 SerAspGlyValAlaIuAlaLeuPheAlaSerGlyHisAlaLeuPheLysValGlyPro 256
 109 GGGCGGGCGCCGACCTGGCGGCGCGC-----GGCGACCGGCATC 147
 257 GlyArgGlyLeuThrThrLeuValArgGlnThrLeuAlaGlyArgGlyAlaAlaIle 276
 148 GCGGCATCGGCATCTGTGCGAGCGCGCGCGCACCGCGGTGCGAGTGGCGGC-- 205
 277 AlaSerLeuGlySer---ThrHisAlaAlaSerGln-ProAlaSerLeuAlaIle 295
 206 -----TGGTGTGCGCGCGCGCTTCATGAC 231
 295 uGlyGlnLeuTrpGluAlaGlyHisAlaValAlaPrtThrAlaArgProArgGly----- 313
 232 TCGCAGACCCGACGACAGACACTACTGCTCAGGCGTGGACATGACCGCCAGATCTCG 291
 314 -ArgProAlaArgSerSerAlaCysArg-----ArgThrArgSer-S 328
 292 CAGGGGTCACACAGGTGTGTCAGGGGCAATTGGGCATCAGCTGGCGCGCGCGCAC 351
 328 eGlyArg-----GlyThrGlySerA 335
 352 GCCAACCCCGCCCGCTGAGCTGTGACGAGGCGGCTTACGTTCCAGCGC 411
 335 rGArgAlaAlaAlaProProProSerArgArgArgProThrAlaArgProSerAlaPro 355
 412 TTGCG-----CGACTACTGACGCGCTTGGCGCGC 441
 355 rGArgTrhProSerArgArgArgArgSerProAlaArgThrProGlyProArgSer--Pr 374
 442 ACGCGCGCGCGCTG-----ACGCGCGCTGATGTGGCGCATTCACGCTG 489
 374 cArgArgArgProSerProSerAlaArgSerProAlaIleTrp----- 388
 490 CGCGCGCGCGCTGATCGCGACTTGCAGCGCGCGCACGAGAGAAATCGCGCGCAG 549
 389 -----GluArgPro-- 391
 550 CGGAGACTGCGCGAGAGACCATGGCAGCGCGCCATCGCATTTGACCGCGCGCTTC 609

392 -----ProArgArgPro-ArgAspArgArgProArg-----ArgLeuLeuA 405
 610 TACCGCGCGCGCGCGCGCGCGCACCGAAGATCATCGAGTGGCGCGCGCTGAC 669
 405 rGProAlaArgArgLeuAla-----ArgArgAspProGlyAspValPro----- 419
 670 GCGCATGGCGCATCTACGCCACCATGCGCGCGCGAGCGAGCGACATCGTGGCGCG 729
 420 -----AsProAlaAlaAlaArgArgAla----- 428
 730 CTGAGAGAACCTTCGCGCATGGCGCGCGAGCTGAGCGTGGCGGTGATCTCGCACAC 789
 429 -----ProAlaAlaGlyAlaLeuProGlyAlaAspLysArg 441
 790 AAGTCATGGCG 836
 441 LysAlaArgGlyProGlyArgArgArgAlaProArgArgGlyAlaAlaAlaArgValPro 461
 837 -----GCTGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 872
 461 rAspArgAlaAspProProArgArgProAlaProAlaValAlaValAlaProAlaVal 481
 873 -----CTGGACGCGTATCGCTACGTCGCGCGCGCGCGCGCGCGCGCGCGCG 912
 481 aValAlaArgSerAlaArgAlaGlyGlnProGlyValGlnHisProValAlaArgLysAla 501
 913 AAGCAGACCG 972
 501 rGAspGlyProAlaArgArgArg-----AlaProProGlnProV 515
 973 GAACTAGCG 1032
 515 aArgAspArgAlaAlaPro-----ArgLysAlaAlaAspArgValGln--ArgG 551
 1033 GTGTGCG 1092
 531 LngArgArgArgAlaValProGly-----ArgArgA 541
 1093 CAGCGCATCTGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1149
 541 LArgGlyProGlyAlaValAlaProAsp-----ValArgProArgAlaProArgLysA 558
 1150 GAGCG 1206
 558 rGlyProArgArgGlyGlyLeuGlyAlaArgProArgGlyGlnProArgAlaValArg 578
 1207 CCGGACCTGGCGCTGCTCCGCTGGAGACGGCGGTATGAAATGACCGCGCTGACCGCG 1266
 578 rGlnProArgAlaAlaProProArgAlaGlyHisProGlnAlaProArgArgAlaArgH 598
 1267 GCGCGCTCG-----CCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1319
 598 LArgAlaArgArgAlaProArgArgValArgArgLeuValGlyArgArgLeuAlaAla 618
 1320 -----CCTGGGTGTTCGACCGCGCGCG 1344
 618 rGAlaArgArgAlaLeuArgArgArgAlaGlyProAlaPheProAlaAlaGlyAlaP 638
 1345 GTGGCGGATACCG 1404
 638 rGlyAlaValAlaArgArgLeu-----ArgArgSerProAlaGly 651
 1405 TACGTCAAGCG 1446
 651 aValAlaArgArgGlyProGlyArgAlaProValLeuAspArgLysAlaLeuGlyA 671
 1447 GCGCGCATGCGCGCGCGCGCGCG-----CGCAGCAGCGCGCGCGCGCGCGCGCG 1497
 671 rGAlaAlaProGlyProArgAlaGlyArgProArgAlaGlyGlyAlaLysAspValAlaA 691
 1498 CAGCCCTTACATCGCGGCGTGAACGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1547

[illegible]

```

QY 879 GCGTATCCCTACGTGGCGGCTCCACCATTGCTCAAGACAGCCGCTGCTGCGCGG 938
DB 322 GYGLIUGLIPROSERGLYCSILLEGHINISERPROGLYALAARG-----GLYARG 339
QY 939 ACCGACCATATATACCTGGTGGCAAGCCCTCCCGAATGAGGCGGCGGACCTGGATGA 998
DB 340 ThrPro-----AspArgGlnSerAlaArgGlyLeu 351
QY 999 AGTCGGCGCGCA---GCGCGGCAATCCAGTACGACGTGGTCCGAGCTGCAGCGCG 1055
DB 352 ProGlnHisArgGlyGlyAspArgLeuGlnProAlaArgPheGlyAlaAspProGlyAlaGly 371
QY 1056 GCGCGG---CATCTACTTCATGATGAGAGCAACCGGACGAGTGCAGCGCATCCT 1103
DB 372 ArgAlaProArgGlyAlaArgThyCysArgSerAspGlyArgGlyGlnCysAlaLeuProSer 391
QY 1104 GCGGT---CGGCGCGCAACCAT-----GAT 1124
DB 392 GYLIHLEHLEGLNPROASPRArgArgSerGlyGlyAlaArgAlaAlaGlySerArgAlaAsp 411
QY 1125 GCGGTCCGACGCGCTGCGCGGACGAGAGCGCGGCGCATCGCGGCTGGGCGACCTTCC 1184
DB 412 ArgAlaGlnHisArgArgAspProAlaArgSerProAlaGlnAlaPro----- 426
QY 1185 GCGGTGCTGGGCGCACTATGCGCGGACCTGGGCTGTTCCGCTGAGAGCGGCGGTATG 1244
DB 427 -----GlyArgAlaArgAlaGlnProGlyThrGlnProGlyAlaGly----- 441
QY 1245 GAGATGACCGGCTGACCGCGCGGCTTGGGCTGCG-----CGGCGCGGCGGCGACT 1298
DB 442 -----GlnAlaAspCysArgArgLeuArgProGlyLeuArgThrProAlaGln 458
QY 1299 GCGAGCGCG---GTACTTCGCGGCACTGTGGTGTGCA---CGGCGGCAAGGTGGC 1339
DB 459 AlaAlaGlyHisProAlaLeuAspArgGlnProAlaGlyAlaHisAspProGly----- 475
QY 1350 CGATACCGCCACCTTCGACACCTACCGA-----GCGCGCGCGCGGCGCAT 1394
DB 476 -----ArgGlnHisArgAlaGlyCysGlnTyLeuGlyGlyLeuArgArgAsp 493
QY 1395 CCAATCCGCT-ACGTCAACGCGCGCGGCTGCGCAAGAGCGGCTTACCGCGGCGAGC 1453
DB 494 ArgLeuArgLeuThrSer-----SerGlyAlaAspArgGlnSerProAlaSer 509
QY 1454 ATCCGCGCGGCTGCTCGGACGAGCGCGGCTGAGCCGCGCGGCGGCGCTTACATCGG 1513
DB 510 -----AlaGlyLeuPheHisGlyArgPro----- 517
QY 1514 GCGTGAAAGCGGCGCGGCTGCGCGGCTTCCCAACCTGGAGCGAAACGCTACATGGCC 1573
DB 518 -----ValGlyAlaArgAla----- 522
QY 1574 CTCCCTCCGCTCGCAATACGCGCCCAACCGATATCGTGGGCAAGGAG 1621
DB 523 LeuProHisAlaValAlaGlyGlyProPheCysArgSerThrPylsLys 538

```

RESULT 5

US-09-413-814-79

Sequence 79, Application US/09413814

Patent No. 6223064

GENERAL INFORMATION:

APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH

APPLICANT: Bristol-Myers Squibb, Co.

APPLICANT: Beyer, Stefan

APPLICANT: Bloeker, Helmut

APPLICANT: Brandt, Petra

APPLICANT: Cino, Paul M

APPLICANT: Dougherty, Brian A

APPLICANT: Goldberg, Steven L

APPLICANT: Hotfel, Gerhard

APPLICANT: Mueller, Joachim

APPLICANT: Reichenbach, Hans

TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413, 814
EARLIER FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 79
LENGTH: 1213
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-79

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1.73e-13	309.00	32.13%	30.77%	9.53%	1213	204	9	250	200	35

US-10-009-782-1 (1-1758) x US-09-413-814-79 (1-1213)

```

QY 1756 GCTTCGTAGCTACAGGCGCATCTGCGCCAGCTCGGCTTGCAGAGGCTCGAGACGCGA 1697
DB 18 AlaArgArgGlyAspArgArgGlyGlyLeuAlaLeuArgSerAspArgGlnAlaArg 37
QY 1696 TCGCGTGGCGCTCGACAGAGTGTGCGAGGCTCAATTTCCGGGCTTGGCTGCGCGGCA 1637
DB 38 ArgArgGlyHisGlyGlnGlyLeuGlyAla-----GlyGlyArgArgAlaArgAla 55
QY 1636 GCGGCGCGCCATCATCTCTTCCGTCGCGGATGCGGCGGCGGCGGATGCGAGCGG--- 1580
DB 56 AlaArgArgArgArgLeuProAlaPro-----GlyProSerProProArgGly 71
QY 1579 -----AGGAGGCGCCATGTAGCGGCTTGGCTCCAGGCTTG 1544
DB 72 AlaProGlnAlaProProGlyAlaArgArgGlyPro-----ProArgProHisAla 88
QY 1543 GAGAGGCGCGGCGACCGCGCGGCTTACCGCGGATGTGAGGCTGCGCGGCGGCTCAGG 1484
DB 89 ValGlyGly-----ArgArgArgArgLeuAlaArgGlyHisArg 102
QY 1483 CGGCGGTGCTCGGACGACGCGCGCG-----CATGCTGCGGCGGCGGCGGCTCT 1433
DB 103 AlaPro-----ArgArgArgArgGlyArgAlaValAlaArgAlaArgAlaArgAla 121
QY 1432 CTGCGCAGACCGCGCGGCTGACGTACAGGATGGATGCGG----- 1368
DB 122 -----AlaArgArgGlyHisGlyArgProGlyLeuArgSerLeuHisValArg 137
QY 1387 -----CGGCGCGCTGCGGCTGAGGCTGTTCGAAGGTGG 1358
DB 138 ValHisGlyAlaAlaGlnGlyArgAspArgArgProArgArgGlyGlnHisGlyPro 157
QY 1357 CGGTATCGGACACCGCTGCGCGGCTGCGCAACACAGAGTGGCGGAGTACCGGCGCTGCA 1298
DB 158 ArgHisGlnProProLeu-----ArgArgArgProGlyGlyProGlyAlaArgAlaLeu 175
QY 1297 GCTGCGCGCGCGCGGCA-----GCGCGAAGCGCGCGGCGGCTCAGCGCGG 1253
DB 176 ValAlaGlnLeuArgProValaGlyLeuArgArgValaArgAspAlaArgArgArgArg 195
QY 1252 TCATCTTCATACCGCGGCTGCA----- 1229
DB 196 ArgArgAspProArgProArgProArgArgLeuGlySerGlyAlaLeuAlaArgAlaArgGly 215
QY 1228 -----GCGGGA-----ACAGGCCAGGCTCG-----GCGCATAGTGGC 1196
DB 216 AlaArgAlaGlyAspArgValaGlnLeuGlyProGlyAlaAspGlyAspAlaHisGlyArg 235

```

[illegible]

```

Db      535  PROLAATG-----ArgLAhIAstArgLAIAArgLAIAArgLAIAArgLAIA 1111111111
QY      235  GCGAGTCGATGAGAGCGGGGGCGGCACACAGGCCGACACGTCGACCCGGGTGTGCGCG 17
Db      553  GLYArgValAlaAspArgProValProValProAspHisArgAlaThrGlyAlaAla 57
QY      175  CGGCGTCGACAGATGCGCGATGCGCGATGCGGTGCGCGCGCGCGACCGCCAGTCGCGCG 11
Db      573  ArgArgGlyAspArgLysLh-----GlyArgLAIAArgLAIAArgLAIAArgLAIA 59
QY      115  GCGGCCCCG 107
Db      591  GLYAlaPro 593

RESULT 6
US-09-219-849-6
/ Sequence 6, Application US/09219849
/ Patent No. 6150081
GENERAL INFORMATION:
APPLICANT: VAN HEERDE, GEORGE V.
APPLICANT: VAN RIJN, ALEXIS C.
APPLICANT: BOWSTRÄ, JAN B.
APPLICANT: DE WOLF, FREDERIK A.
APPLICANT: MOORROEK, ANDREAS
APPLICANT: WERTEN, MARC W.T.
APPLICANT: WIND, RICHELIE D.
APPLICANT: VAN DEN BOSCH, TANJA J.
TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
FILE REFERENCE: 2728-2
CURRENT APPLICATION NUMBER: US/09/219, 849
CURRENT FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 960
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Illustrative
US-09-219-849-6

Alignment Scores:
Pred. No.: 4.74e-12 Length: 960
Score: 287.50 Matches: 196
Percent Similarity: 35.45% Conservative: 33
Best Local Similarity: 30.34% Mismatches: 262
Query Match: 8.71% Indels: 157
DB: 4 Gaps: 40

US-10-009-782-1 (1-1758) x US-09-219-849-6 (1-960)
QY      63  CCGTGCTGCTCGGCGGCGGCAACCCATCATGAGCGGACAGACACCCCGGGGGCGGCCGA 122
Db      205  ProGlyLAIAgLnGly-----ProLAIAIAProGlyGlySerLAIAgAsp 218
QY      123  CCGTGGCGGTGCGGCGGCGACCGCATGCGCGGCATGCGGATCTGTGGAACGCGCGCGCA 182
Db      219  ProGly-----ProProGlyLAIAHis---GlyProLAIAIAArgLAIAArgLAIA 233
QY      183  CAC-----CGGGGTCGACGTGTGCGGCGCGGTGTGCGCGCGCGCGCTTATCGA 230
Db      234  HisGlyProLAIAIAArgLAIAArgLAIAHisGlyProLAIAIAArgLAIAArgLAIA 253
QY      231  CTCGACACACCACAGA---CGACAACTACTGCT---CAGGCGTCGCGACATGACGCCAA 284
Db      254  ProLAIAIAArgLAIAArgLAIAIAgLnGlyProLAIAIAArgLAIAArgLAIAArgLAIA 273
QY      285  GATGTCCAGAGGGGTGCACACAGGTGTGTGTCAGGGCAATTGGCGCATACGCTTGCGCGCT 344

```

```

OY 1266 GCGGCTTGCGCTGGCCGGGCGCGGAGACAGCTCAGAGCCGGGTACTTCGGCCAGCTGGTGG 1327
Db 617 -----ProGly-----AlaGlnGlyProIaGly-----ProGlyG1 627
OY 1328 TGTTCGA-----CCGGGCGCA-CGGTGGCCGATACCGCCAGCTTCGACACCTT 1374
Db 627 ySerAArgspProGlyProProGlyAlaGlnGlyProIaGlyProGlyGlySerAArg 647
OY 1375 ACCGAGCGCGCGCGCGCGCATCCGCTGTAAGTCAGCAACGGCGCGCGCGTGTGGCAAG 1434
Db 647 pProGlyProProGlyAlaGlnGlyPro-----AlaGlyProGlyGlySerA 663
OY 1435 CAGGCGCTTCACG-----GCCAGCATGCGCGCGCGCTGCTCGCAGCAGCGCGCGCT- 1486
Db 663 gAspProGlyProProGlyAlaGlnGlyProIaGlyProGlyGlySerAArgspProG1 683
OY 1487 -GAGCCCGGCGCGCAGCCCTTACATCCGGCGGTGAACGGCGCGCGCTGCCCGCTCCCA 1545
Db 683 yProProGlyAlaGln-----GlyProAla-GlyProGlyGlySerAArgspProGlyP 701
OY 1546 ACCCTGGAGCCAAACCGCTCATGCGCCCTCC--CTCGCGCTGGCAATACGCCCCACCC 1602
Db 701 roProGly-----AlaHisGlyProIaGlyProIaGlyAlaHisGlyProIaG 718
OY 1603 GAATGCTGGG-----CAGGAAGATGATGGCGCGCGCCCTGGCGC 1641
Db 718 yProIaGlyAlaHisGlyProIaGlyProIaGlyAlaHisGlyProIaGlyProI 738
OY 1642 GCCGAGCGCAAGGCCCG-----GAAATGACCCCTGCAGACCTGTGCGCA- 1685
Db 738 yGcGlyAlaGlnGlyProIaGlyProGlyGlySerAArgspProGlyProProGlyAlaG 758
OY 1686 -----GCCAGCGCGCATCGGGGCTCGACCCCTGTCCAGGCGCGAG 1725
Db 758 lngIyProIaGlyProGlyGlySerAArgspProGlyProProGlyAlaGlnGlyProA 778
OY 1726 CTGGGCCCA 1733
Db 778 lngIyPro 780

```

RESULT 7
US-07-757-022B-14
Sequence 14, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating F
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cseri, Luan
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 941 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-14

Alignment Scores:
 Pred. No.: 6,45E-12 Length: 941
 Score: 285.50 Matches: 137
 Percent Similarity: 33.50% Conservative: 60
 Best Local Similarity: 23.30% Mismatches: 240
 Query Match: 8.65% Indels: 151
 DB: 4 Gaps: 24

US-10-009-782-1 (1-1758) x US-07-757-022B-14 (1-941)
 QY 32 CCATGCCCAATCCGATTCACGACCCCTGCTGCTGCGGGGCGACCCCTACG 91
 Db 167 ProThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThr 186
 QY 92 ACAGGACCA-----ACACCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 142
 Db 187 ThrLysSerAlaProThrThrProLysGluProAlaProThr-----Thr 201
 QY 143 GCATGCCCGCCATCGGAGATCTGCGACGCGCGCGACACCCGGGTGACGTGCG 202
 Db 202 ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProThr 216
 QY 203 GCGTGTGTCGCGCGCGCGCTCATGCTGCGACGCCACCGACGCACTACTGCTCA 262
 Db 217 -----ThrThrLysGluProAlaProThrThrThrThrLysSer 228
 QY 263 GGGGTGCGACATGAGCGCCAGATCTCGCAGGGCGCTCACACGCTGCTCAGCGCAT 322
 Db 229 AlaProThrThr-----ProLysGluProAlaPro----- 238
 QY 323 GCGGATCAGCTGCGCGCGCTGCGCAGCCACACCCCGCGCGCGCTGACCTGCTG 382
 Db 239 -----ThrThrProLysLysProAlaProThrThrProLys 250
 QY 383 ACAGAGCGGCTCTACCGTTTCGAGCGGCTTCGCGACTGACGCGGTGCGGCGCA 442
 Db 251 GluProAlaProThrThrProLysGluProThrThrThrProLysLysProAlaPro 270
 QY 443 CGCGGCGG-----CCCTCAACGCGCGCTGTATGGTGGCCATTGACGCTGCGCG 493
 Db 271 ThrThrLysGluProAlaProThrThrPro----- 280
 QY 494 CGCGGTCATCGCGGACTGACGCGCGCGCGCGACGAGAGAAATCGCGGCGATCGCGG 553
 Db 281 -----LysGluProAlaProThrThrAlaProLysLysProAlaPro 293
 QY 554 ACCTGGCGGAGAGACATGCGCGCGCGCGCGCGCGCGCGCGCGCTCTACG 613
 Db 294 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 311
 QY 614 CGCGCGCGCGCGCGCGCGCA-----CCACCGAGAGATCATGAGGTGCGCGCGCGTGA 670

Db 312 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrThrLysSerAla 331
 QY 671 CGCATGGGCGGATACGCCACCCACATGCGCGAGAGAGCGGACACATGTCGCGCGCG 730
 Db 332 ProThrThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 351
 QY 731 TGGAGAAACCTCCGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 790
 Db 352 GluProSerProThr-----ThrThrLysGluProAlaProThrThr 365
 QY 791 AGGTCATGGCG 850
 Db 366 ProLysGluProAlaPro-----ThrThrPro 374
 QY 851 CCATGGCG 910
 Db 375 LysLysProAlaProThrThrThrProLysGluProAlaProThrThrThrProLysGluProAla 394
 QY 911 TCAAGCAGAGACCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
 Db 395 ProThrThrThrThrLysLysProAlaProThrThrAlaProLysGluProAlaProThrPro 414
 QY 971 CCGAATGAGCGGCG 1030
 Db 415 LysGluThrAlaProThrThrProLysLysLeuThrProThrThrProLysLysLeuAla 434
 QY 1031 ACGGTGTCG 1090
 Db 435 ProThrThrPro-----GluLysProAlaProThrThrProLysLysLeuAlaProThr 452
 QY 1091 TCGACGCGATCTGCG 1147
 Db 453 ThrProGluLysProThrProThrProThrProGluLysProAlaProThrThr 469
 QY 1148 ACAGCG 1207
 Db 469 ----- 469
 QY 1208 GCGACCTGGCG 1267
 Db 470 -----Pro 470
 QY 1268 CGGCTTCG 1327
 Db 471 LysAlaAlaAla-----ProSerThrProLysGluProAlaProThrThrProLysGluPro 489
 QY 1328 TGTTCGACCG 1384
 Db 490 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 509
 QY 1385 CGCGCGCATTCATTCGCTGACGTCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1444
 Db 510 ProLysGluThrAlaPro-----ThrThrLysGluProAlaProThrThrProLysLys 528
 QY 1445 CGCGCGCATCCG 1504
 Db 529 ProAla-----ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr 545
 QY 1505 TACAAATCGCGGCGTAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1564
 Db 546 -----ThrSerAspLysProAlaProThrThrProLysGluThrAlaPro 560
 QY 1565 ACATGGCCCTCCCTCCCTCGCATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624
 Db 561 ThrThrProLysGluProAlaProThrThrPro----- 571
 QY 1625 TGGGCG 1684
 Db 572 -----LysGluProAlaProThrThrProLysGluThrAlaPro-----ThrThrLys 588
 QY 1685 AGCGCAGCG-----GCATGCGCGGTCTCGACCGCTGTGCGAGG-----CCGAGCTGGCG 1732


```

QY 1148 ACGAGCGCCGCGCATTCCGGGCGCTGTGGGACACTTCCCGGGGTGCTGGGGCACTATTCGCC 1207
Db 550 ----- 550
QY 1208 GCGACCTGGGCTGTTCCTCCCTGGAGAGCGGGGATGAGAAATGACGGCGCTGACCGCGC 1267
Db 551 -----Pro 551
QY 1268 CGCGCTTCGGCTGCGCGGGCGGGGAGCTGACGCGCGGTACTTCGCGCAGCTGTGG 1327
Db 552 LysAlaAlaAla-----ProAsnThrProLysIubProAlaProThrThrProLysIubPro 570
QY 1328 TGTTCGACCGCGCGAGGATGGCCGATACCG-----CCACCTTGAACACCCATACCGAGCGC 1384
Db 571 AlaProThrThrProLysIubProAlaProThrThrProLysIubProAlaProThrThr 590
QY 1385 CGCGCGGCATCCATTCGCTAGCTCAACGCGCGCGCGGTGTGCAAGACGAGCGTTCA 1444
Db 591 ProLysIubProAlaPro-----ThrThrLeuLysIubProAlaProThrThrProLysLys 609
QY 1445 CGCGCCGACGATGCCCGGCGCGGTGCTGCAACGACGCGCGCGCTGAGCCGCGCGCGCCCT 1504
Db 610 ProAla-----ProLysIubLeuAlaProThrThrThrLysIubProThrSerThr--- 626
QY 1505 TACAAATCCGGCGTGAACGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1564
Db 627 -----ThrSerAlaLysProAlaProThrThrProLysIubThrAlaPro 641
QY 1565 ACATGCGCCCTCCCTCCCTCGCGCGCAATACGCGCCCGCCCGCATTCGTGGGCAAGAACTGA 1624
Db 642 ThrThrProLysIubProAlaProThrThrPro----- 652
QY 1625 TGGGCGCGCGCTGCGCGCGCGAGCGCAAGCGCGAAATGACCTGCAAGACCTGTGCG 1684
Db 653 -----LysGluProAlaProThrThrProLysIubThrAlaPro-----ThrThrLeuLys 669
QY 1685 AGGCGACGCG-----GCATCGCGGCTGTGCAACCTGTGCAAG-----CCGACCTGGGCGC 1732
Db 670 GluProAlaProThrThrProLysIubProAlaProLysIubLeuAlaProThrThrThr 689
QY 1733 AGATGCGCCCTGAGCTACGAGAAAGC 1756
Db 690 LysGluProThrSerThrThrSer 697

RESULT 9
US-07-757-022B-74
; Sequence 74, Application US/07757022B
; Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502

```

```

? FILING DATE: 18-JAN-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/546,114
? FILING DATE: 29-JUN-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/457,196
? FILING DATE: 29-DEC-1989
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/390,901
? FILING DATE: 08-AUG-1989
? ATTORNEY/AGENT INFORMATION:
? NAME: Cseri, Luanan
? REGISTRATION NUMBER: 31,822
? REFERENCE/DOCKET NUMBER: GI 5190
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617)876-1170
? TELEFAX: (617)876-5851
? INFORMATION FOR SEQ ID NO: 74:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1038 amino acids
? TYPE: AMINO ACID
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-07-757-022B-74

Alignment Scores:
Pred. No.: 6,61e-12 Length: 1038
Score: 285.50 Matches: 137
Percent Similarity: 33.50% Conservative: 60
Best Local Similarity: 23.30% Mismatches: 240
Query Match: 8.65% Indels: 151
DB: 4 Gaps: 24

US-10-009-782-1 (1-1758) x US-07-757-022B-74 (1-1038)
QY 32 CCAATGCCAAATCCGATTCCTCCAGCCCTTGAGACCTGCTGCTCGGGGCGGACCCATCATCG 91
Db 232 ProthPrProthThrThrIleYsSerAlaProThrThrProLYsGluProAlaProThrThr 25
QY 92 ACGGCAGCA-----ACACC CGGGGGGCGGCGCGGACCGACCTGGGCGTGCGGCGAGC 14
Db 252 ThLYsSerAlaProThrThrProLYsGluProAlaProThr-----Thr 26
QY 143 GCATGCCCGCCCATCGGCGATCTGTGGAGCGCCGCGCGACACCCGGGTGCAGCTGTGCG 20
Db 267 ThrLYsGluProAlaPro-----ThrThrProLYsGluProAlaProThr----- 28
QY 203 GCGTGTGGTGC GCGCCGCGCTCATCGACTGCACACCCAGAGCAACAACACTGCTCA 26
Db 282 -----ThrThrLYsGluProAlaProThrThrThrLYsSer 29
QY 263 GCGCGTCGCACTGACGCCCAAGATCTCGCAGGGGCGTCACACGCGTGATCAGCGCAATT 32
Db 294 AlaProThrThr-----ProLYsGluProAlaPro----- 30
QY 323 GCGGCAATGAGCCTGGCGCGGCGTGCGGCGACCGCAACCCGCCGCCCTTGAGACTGCTCG 38
Db 304 -----ThrThrProLYsSPAlaProThrThrProLYs 31
QY 363 ACGAAGCGGCTCTTACGTTTCGAGCGGCTTGC GCGCACTACCTGAGCGCGTGGCGGCCA 44
Db 316 GluProAlaProThrThrProLYsGluProThrProThrThrProLYsGluProAlaPro 33
QY 443 CGCGGGGG-----CGTCACAGCCCGCGCTGTATGTGGGCATTCACGCTCGCGG 49
Db 336 ThrThrLYsGluProAlaProThrThrPro----- 34
QY 494 CCGGCGTATCGCCGAGCTTGCGAGCGCCCGCCACCGACGAGAAATCGCGGCATCGGG 53
Db 346 -----LYsGluProAlaProThrThrAlaProLYsSPAlaPro 35
QY 554 ACCTCGCCGAGGAACCATATGCGCCACGCGCGCCATCGGACTTTCGACCGCGCGCTTAC 61
Db 554 -----LYsGluProAlaProThrThrAlaProLYsSPAlaPro 35

```

```

Db 359 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 376
QY 614 GCGCCGCGCGCGCGCGCA---CCAGCGAGAGATCATGAGAGTGTGCGCGCGGTGAGCG 670
Db 377 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrThrLysSerAla 396
QY 671 CGCATGGCGGCGCATGACGCGCGAGCATGCGCGAGCGAGCGAGCATGCGCGCGCG 730
Db 397 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 416
QY 731 TGGAGAAACCTTCGCGCATGCGCGAGCGAGCGAGCGCGGCGGTATCTCCACACACA 790
Db 417 GluProSerProThr-----ThrLysGluProAlaProThrThr 430
QY 791 AGGTATGGGCGCGAGCGCATTTGGCGCGCTGCGCGAGACGTCGCGCGGTATGAGAGCGG 850
Db 431 ProLysGluProAlaPro-----ThrThrPro 439
QY 851 CCATGGGCGCGCGAGCGTCTGCGAGCGGATACCTACGTGCGCGCGGTCCACCATGCG 910
Db 440 LysLysProAlaProThrThrThrProLysGluProAlaProThrThrProLysGluProAla 459
QY 911 TCAGAGAGAGCGCGCGTGTGCGCGAGCGAGCATCATCACTGTCGTCGACGCGCTTCC 970
Db 460 ProThrThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 479
QY 971 CCGAAGTACGAGCGCGCGAGCGATGAGATGCGCGCGCGCGCGCAATCCAGTAGCG 1030
Db 480 LysGluThrAlaProThrThrProLysLysLeuThrProThrThrProGluLysLeuAla 499
QY 1031 ACGTGTCGCGCGAGTGTGAGCGCGCGCGCGCATCTACTATGATGAGACGACCGCGAG 1090
Db 500 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLeuAlaProThr 517
QY 1091 TGCAGCGCATCTGCGGTGCGCGCGCGAGCATGATCG---GCTCGGACGCGCTGCGCGAG 1147
Db 518 ThrProGluLysProThrProThrThrProGluLysProAlaProThrThr----- 534
QY 1148 ACGAGCGCGCGCATCGCGCTGTGGGCGCACCTTCCGCGCGGTGCTGGGCGCATATGCGC 1207
Db 534 ----- 534
QY 1208 GCGACCTGCGGCTGTCCCGCTGTGAGAGCGCGGTATGAGATGACCGCGCTGACCGCG 1267
Db 535 ----- 535
QY 1268 GCGCTTTCGCGCTGCGCGCGCGCGCGAGCTGAGCGCGCGGTACTTCCGCGACCTGTGG 1327
Db 536 LysAlaAlaAla---ProAlaThrProLysGluProAlaProThrThrProLysGluPro 554
QY 1328 TGTTCGACCGCGCGCGAGTGTGCGATACG---CCAGCTTGAACACCTTACCGAGCGCG 1384
Db 555 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 574
QY 1385 CCGCGCGCGCATTCGCTGCTAGTCAAGCGCGCGCGGTGCGCAAGAGAGCGGTCA 1444
Db 575 ProLysGluThrAlaPro---ThrThrLysGluProAlaProThrThrProLysLys 593
QY 1445 CCGGCGAGCATGCGCGCGCGTGTGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCT 1504
Db 594 ProAla-----ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr 610
QY 1505 TACATCCGCGCGTGAACGGGCGCGGTGCGCGCGCTCCGCGACCGTGGAGCGAACCCT 1564
Db 611 ----- 611
QY 1565 ACATGCGCGCGCTTCGCGTGCATATGAGCGCGCGCGCATATGCTGGGCAAGAGATGA 1624
Db 626 ThrThrProLysGluProAlaProThrThrPro----- 636
QY 1625 TGGGCGCGCGCTGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1684
Db 637 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLysLys 653

```

```

QY 1685 AGCGCAGCG---CGATCGCGGCTGCGACCTGTGCAAG-----CCGAGCTGGCGC 1732
Db 654 GluProAlaProThrThrProLysLysProAlaProLysGluLeuAlaProThrThrThr 673
QY 1733 AGATCGCGCTGAGCTACGAGAGCG 1756
Db 674 LysGluProThrSerThrThrSer 681

RESULT 10
US-07-757-022B-58
; Sequence 58, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseiff, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-58

Alignment Scores:
Pred. No.: 6.63e-12 Length: 1049
Score: 285.50 Matches: 137
Percent Similarity: 33.50 Conservative: 60
Best Local Similarity: 23.30 Mismatches: 240
Query Match: 8.65e Indels: 151
DB: 4 Gaps: 24

US-10-009-782-1 (1-1758) x US-07-757-022B-58 (1-1049)

```

```

Db      561  ThProgluInuProthrProthrProthrProgluInuProAlaProthrThr----- 577
Oy      1148  ACGAGGCGCCGCATCCGGCCCTGTGGGACCCCTCCGCGGGTCTGGGCGACTATGCGC 1207
Db      577  -----
Oy      1208  GCGACCTGGGCGTGTCCCGCTGAGAGCGGGGTATGGAAGATGACCGGCTATCCGCCG 1267
Db      578  -----Pro 578
Oy      1268  CGGCTTCGGCTGGCGCGGGCGGGGAGCTGACAGCCGGGTACTTCGCCACCTGTGTG 1327
Db      579  LysAlaAlaAlaAla---ProAsnThrProLysgluProAlaProthrThrProLysgluPro 597
Oy      1328  TGTGTGACCCGCGCACGGGGGCGGATACCG---CCACCTTCGAAACACCTACCGAGCGG 1388
Db      598  AlaProthrThrProLysgluProAlaProthrThrProLysgluThrAlaProthrThr 617
Oy      1385  CGCGCGGACATCCATTCGCTGTACGTCAACGGCGCGCGCTGTGGCAAGAGCGGGGTCA 1444
Db      618  ProLysgluThrAlaPro---ThrThrLeuLysgluProAlaProthrThrProLysLys 636
Oy      1445  CGGGCGAGATGCCGGCGCGCTGCTCCGACGACAGCGCGCTGAGCGCGCGCCAGCCCT 1504
Db      637  ProAla-----ProLysgluLeuAlaProThrThrThrLysgluProthrSerThr--- 653
Oy      1505  TACAATCCGGCGTGAAGAGGGGCGGCGAGGCGCCGCCCTCCCAACCTCGAGACGAACCGCT 1564
Db      654  -----ThrSerAspLysProAlaProthrThrProLysgluThrAlaPro 668
Oy      1565  ACATGCCCCCTCCCTCCGCTCGCATATAGGCCGCCACCGCATATCGTGGGCAAGATGA 1624
Db      669  ThrThrProLysgluProAlaProthrThrPro----- 679
Oy      1625  TGGGCGCGCGCTCGGGGCGGCGGACGCAAGGCCCGGAAAAATGACCTGCAGACCTGTGC 1684
Db      680  -----LysgluProAlaProthrThrProLysgluThrAlaPro---ThrThrLeuLys 696
Oy      1685  AGCGACGCG---GCATGCGCGGTCTGACCCCTGTCCAAAG-----CCGAGCGTGGGCC 1732
Db      697  GluProAlaProthrThrProLysLysProAlaProLysgluLeuAlaProthrThrThr 716
Oy      1733  AGATCGCCCTGAGCTATCGAAGAC 1756
Db      717  LysgluProthrSerThrThrSer 724

RESULT 11
US-07-757-0228-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910

```

Pred. No.:	6, 77e-12	Length:	1140
Score:	285.50	Matches:	137
Percent Similarity:	33.50%	Conservative:	60
Best Local Similarity:	23.30%	Mismatches:	240
Query Match:	8.65%	Indels:	151
DB:	4	Gaps:	24

US-10-009-782-1 (1-1758) x US-07-757-022B-104 (1-1140)

[illegible][illegible]

QY 1625 TGGGCGCGCGCTGCGCGCGAGCGAGCGCCGGAATGACCTGCAAGACCTGTGCG 1684
 Db 771 -----LysGluProAlaProThrThrProLysGlyThrAlaPro---ThrThrLeuLys 787
 QY 1665 AGCCAGCG--GCATCGCGGTCTGACCGCTGCAAGG-----CCGAGCTGGCGC 1732
 Db 788 GluProAlaProThrThrProLysProAlaProLysGluLeuAlaProThrThrThr 807
 QY 1733 AGATCGCCCTGAGTACAGAGAC 1756
 Db 808 LysGlyProThrSerThrThrSer 815
 RESULT 12
 US-07-757-022B-44
 Sequence 44, Application US/0757022B
 Patent No. 6433142
 GENERAL INFORMATION:
 APPLICANT: Gesner, Thomas G.
 APPLICANT: Clark, Stephen C.
 APPLICANT: Turner, Katherine
 APPLICANT: Hewick, Rodney M.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS: 143
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/757,022B
 FILING DATE: 19910910
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Geert, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 44:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1270 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-44
 Alignment Scores:
 Pred. No.: 6.95e-12 Length: 1270
 Score: 285.50 Matches: 137
 Percent Similarity: 33.50% Conservative: 60
 Best Local Similarity: 23.30% Mismatches: 240
 Query Match: 8.65% Indels: 151
 Db: 4 Gaps: 24

US-10-009-782-1 (1-1758) x US-07-757-022B-44 (1-1270)
 QY 32 CCATGTCCCAATCCGATTCACAGCCCTTGCAGCTGTCTGCGGGGCGACCCCTCATCG 91
 Db 232 ProThrProThrThrThrLeuLysSerAlaProThrThrProLysGluProAlaProThrThr 251
 QY 92 AGCGCAGCA-----ACACCCCGGGGGCGCGCGCGGACCTGGCGCGCGCAGC 142
 Db 252 ThrLysSerAlaProThrThrProLysGluProAlaProThr-----Thr 266
 QY 143 GCATCGCGCATGCGCGCATCTGTGAGCGCGCGCGCGCACACCCGGGTGAGCTGTGCG 202
 Db 267 ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProThr----- 281
 QY 203 GCGTGTGTGTGCGCGCGCGCGCTTCACTGACGCGACCCAGCAGCAGCACTACTGTCTCA 262
 Db 282 -----ThrThrLysGluProAlaProThrThrThrLysSer 293
 QY 263 GCGCTGCGCATGACGCGCCCAAGATCTGCGAGGGCGGTACACGAGTGTACGCGGCAATT 322
 Db 294 AlaProThrThr-----ProLysGluProAlaPro----- 303
 QY 323 GCGGATCAGCTGCGCGCGCGCTGCGCGCAACCGCGCGCGCGCTGAGCTGTG 382
 Db 304 -----ThrThrProLysSerProAlaProThrThrProLys 315
 QY 383 ACAGAGCGCGCTTACCGTTTGAGCGCTTCCGCGCATCTGACGCGGTGCGGCGCA 442
 Db 316 GluProAlaProThrThrThrProLysGluProThrThrThrProLysGluProAlaPro 335
 QY 443 CGCGCGCGG-----CGGTACAGCGCGCGGTATGATGAGGCGCATTCACGCGCGG 493
 Db 336 ThrThrLysGluProAlaProThrThrPro----- 345
 QY 494 CGCGGTCATGCGGACTGTGACAGCGCGCGCGCACGAGAGAAATCGGCGCATGCGG 553
 Db 346 -----LysGluProAlaProThrAlaProLysProAlaPro 358
 QY 554 ACCTGCGCGAGAGACCATGAGCGCGCGCATGCGCATTTGACGCGCGCTCTAC 613
 Db 359 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 376
 QY 614 CGCGCGCGCGCGCGCA---CCACGGAAGATCATCGAGGTGTGCGCGCGCTGACG 670
 Db 377 ThrLysGluProSerProThrThrThrProLysGluProAlaProThrThrThrLysSerAla 396
 QY 671 CGCATGCGCGCATCTACGCGCACCATGCGCGAGAGAGCGACATCTGTGCGCGC 730
 Db 397 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 416
 QY 731 TGGAGAAACCTTCGCGCATGCGCGCGAGCTGACGTCGCGGTGTGATCTGCGACCA 790
 Db 417 GluProSerProThr-----ThrThrLysGluProAlaProThrThr 430
 QY 791 AGTATGCGCGCAGCCCAATTTCGCGCGCGCTGCGGAGACCGCTGCTGATCGAGCGC 850
 Db 431 ProLysGluProAlaPro-----ThrThrPro 439
 QY 851 CCATGCGCGCGAGACGCTCTGCTGAGCGGTATCCCTACGTGCGCGCGCTCCACCATTC 910
 Db 440 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 459
 QY 911 TCAAGCAGACCGCGGTGTGCTGCGCGAGCAGCATCATCATCTGTGACAGCCCTTC 970
 Db 460 ProThrThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 479
 QY 971 CCGAATGAGCGCGCGGACCTGATGATGAGCGCGCGAGCGCGGCAATTCAGTACG 1030
 Db 480 LysGluThrAlaProThrThrProLysLeuThrProThrThrProLysLeuAla 499
 QY 1031 AGTGTGCGCGAGCTCAGCGCGCGCGCATCTACTCATGATGAGAGACCGGAGC 1090
 Db 1031 -----

Db 500 ProThrPro-----gluysProAlaProThrThrProgluLeuAlaProThr 517
 QY 1091 TGCAGCCATCTGGCGCTGGCCGACATGATCG---GCTCCGACGCCCTGCCAGC 1147
 Db 518 ThrProgluLeuProThrProThrProThrProgluLeuAlaProThr----- 534
 QY 1148 ACAGAGCCCGCATCCGCGCTGTGGGACCTTCCCGGGGTGTGGGGCACTATGCGC 1207
 Db 534 ----- 534
 QY 1208 GCGACCTGGGCTGTCCGCTGGAGACGGGTATGAGATGACGGCGCTGACGCCG 1267
 Db 535 -----Pro 535
 QY 1268 CGCGCTTCGGCTGGCGCGCGGCGACCTGACAGCGCGGTACTTCCGCGACCTGGTG 1327
 Db 536 LysAlaAlaAla---ProsnThrProLysgluProAlaProThrProLysgluPro 554
 QY 1328 TGTGACCGCGGACAGCGGTGGCGCATACCG---CCACCTTCGACACCTTACCGAGCGG 1384
 Db 555 AlaProThrThrProLysgluProAlaProThrThrProLysgluAlaProThrThr 574
 QY 1385 CGCGCGCATCCATCCGTCGTACGTCAGAGCGCGCGCTGTGGCAAGACGAGCGCTCA 1444
 Db 575 ProLysgluThrAlaPro---ThrThrLeuLysgluProAlaProThrThrProLys 593
 QY 1445 CGCGCGCATCCGCGCGCGCTGTGTCGACGACGCGCGCGCTGAGCGCGCGCGCGCT 1504
 Db 594 ProAla---ProLysgluLeuAlaProThrThrThrLysgluProThrSerThr--- 610
 QY 1505 TACATCCGCGGTGAGAGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1564
 Db 611 -----ThrSerAspLysProAlaProThrThrThrProLysgluThrAlaPro 625
 QY 1565 ACATGCGCCCTCCCTCGCTGCGCATACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624
 Db 636 ThrThrProLysgluProAlaProThrProThrPro----- 636
 QY 1635 TGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1684
 Db 637 -----LysgluProAlaProThrThrProLysgluThrAlaPro---ThrThrLeuLys 653
 QY 1685 AGCGCAGCG---GCATCGCGGTGTCGACCTGTCGACAGG-----CGGACTGGCGC 1732
 Db 654 GluProAlaProThrThrProLysgluProAlaProLysgluLeuAlaProThrThr 673
 QY 1733 AGATCGCGCGGAGCTAGAGAGAC 1756
 Db 674 LysgluProThrSerThrThrSer 681

RESULT 13
 US-07-757-022B-42
 Sequence 42, Application US/07757022B
 Patent No. 6433142
 GENERAL INFORMATION:
 APPLICANT: Gesner, Thomas G.
 APPLICANT: Clark, Stephen C.
 APPLICANT: Turner, Katherine
 APPLICANT: Hewick, Rodney M.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140.
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/757,022B
 FILING DATE: 19910910
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Gseert, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 42:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1311 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-42

Alignment Scores:
 Pred. No.: 7.01e-12
 Score: 285.50
 Percent Similarity: 33.508
 Best Local Similarity: 23.308
 Query Match: 8.658
 DB: 4
 Gaps: 24

US-10-009-782-1 (1-1758) x US-07-757-022B-42 (1-1311)
 QY 32 CCATGTCACCAATCCGATCCGACCCCTGACCTGCTGCGGGCGGACCTCATCG 91
 Db 273 ProThrProThrThrThrLysSerAlaProThrThrProLysgluProAlaProThr 292
 QY 92 ACAGCAGCA-----ACACCCCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 142
 Db 293 ThrLysSerAlaProThrThrProLysgluProAlaProThr-----Thr 307
 QY 143 GCATGCGCGCGCATGCGCATGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 202
 Db 308 ThrLysgluProAlaPro-----ThrThrProLysgluProAlaProThr----- 322
 QY 203 GCGTGGTGTGCG 262
 Db 323 -----ThrThrLysgluProAlaProThrThrLysSer 334
 QY 263 GCGCTCCGACATACG 332
 Db 335 AlaProThrThr-----ProLysgluProAlaPro----- 344
 QY 323 GCGGATCAGCTGCG 382
 Db 345 -----ThrThrProLysgluProAlaProThrThrProLys 356
 QY 383 ACAGAGCGCGCTTACCGTTGAGGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCG 442
 Db 357 GluProAlaProThrThrProLysgluProThrProThrThrProLysgluProAlaPro 376
 QY 443 CGCGCGCGG-----CGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493
 Db 377 ThrThrLysgluProAlaProThrThrPro----- 386

QY 494 CCGGGGTCAATGCGGAGCTTGCAGCGCGCCGACGAGGAAATCCGGCCATGCGG 553
 Db 387 -----LysGluProAlaProThrAlaProLysProAlaPro 399
 QY 554 ACCTGGCCGAGAGCAATGCGGAGCGCGCCATTCGACCGGCGCTTTCAC 613
 Db 400 ThrThrProLysGluProAlaProThrProLys-----GluProAlaProThrThr 417
 QY 614 CCGCCCGCGCGCGCA---CCAGGAGAGATCATCGAGTGTCCGCGCGCTAGG 670
 Db 418 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrLysSerAla 437
 QY 671 CCGATGGCGGCACTACGACCAATGCGGAGCGGAGGAGGAGCATGCGCGCGC 730
 Db 438 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrProLys 457
 QY 731 TGGAGGAAACCTTCGCGATCGCGCGCGAGCTGAGCGGTGATCGACACACA 790
 Db 458 GluProSerProThr-----ThrThrLysGluProAlaProThrThrThr 471
 QY 791 AGGTCAATGCGGCGCAATTCGCGCGCTGCGGAGAGCGTGCCTGATCGAGCGC 850
 Db 472 ProLysGluProAlaPro-----ThrThrPro 480
 QY 851 CCATGGCGCGCGAGAGCTCTGCGTGCAGCGCTATCCATGCGCGCGCGCTCCACATGC 910
 Db 481 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 500
 QY 911 TCAGCAGAGACCGCGCTGCTGCGCGGAGCAGCAGCATATCATCGTGGACGCCCTTCC 970
 Db 501 ProThrThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrPro 520
 QY 971 CCGAATGAGCGGCGCGCGCGCGGAGTGAAGTTCGCGCGCGCGGCAATCCAAATGACG 1030
 Db 521 LysGluThrAlaProThrThrProLysLysLeuThrProThrProLysLysLeuAla 540
 QY 1031 ACGTGGCGCGCGAGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1090
 Db 541 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLeuAlaProThr 558
 QY 1091 TGCAGCGCATCTGCGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1147
 Db 559 ThrProGluLysProThrProThrProThrProGluLysLeuAlaProThr----- 575
 QY 1148 ACGAGCG 1207
 Db 575 ----- 575
 QY 1208 GCGACTGGCGCTGCTCCGCTGAGAGCGCGGTATGAGATGACCGCGCTGACCGCG 1267
 Db 576 -----Pro 576
 QY 1268 CCGGCTTCG 1327
 Db 577 LysAlaAlaAla---ProAlaThrProLysGluProAlaProThrThrProLysGluPro 595
 QY 1328 TGTTCAGCG 1384
 Db 596 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 615
 QY 1385 CCGCGCGCATTCGCGGTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1444
 Db 616 ProLysGluThrAlaPro---ThrThrLysGluProAlaProThrThrProLysLys 634
 QY 1445 CCGGCGCGAGTGGCG 1504
 Db 635 ProAla-----ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr 651
 QY 1505 TACAAATCCGCGGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1564
 Db 652 -----ThrSerLysProAlaProThrThrProLysGluThrAlaPro 666
 QY 1565 ACATGGCGCGCTCCCTCCGCTCGCAATAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624

Db 667 ThrThrProLysGluProAlaProThrThrPro----- 677
 QY 1625 TGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1684
 Db 678 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLys 694
 QY 1685 AGCGCGCG---GCATCGCGGTCTGACCGCTGTCCAGG-----CCGAGCTGGCG 1732
 Db 695 GluProAlaProThrThrProLysLysProAlaProLysGluLeuAlaProThrThrThr 714
 QY 1733 AGATCGCGCTGAGCTAGAGAGAC 1756
 Db 715 LysGluProThrSerThrThrSer 722
 RESULT 14
 US-07-757-022B-142
 Sequence 142, Application US/07757022B
 Patent No. 6433142
 GENERAL INFORMATION:
 APPLICANT: Gesner, Thomas G.
 APPLICANT: Clark, Stephen C.
 APPLICANT: Turner, Katherine
 APPLICANT: Hewick, Rodney M.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/757,022B
 FILING DATE: 19910910
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Casert, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 142:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1313 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-142
 Alignment Scores:
 Score: 7.01e-12 Length: 1313
 Pred. No.: 285.50 Matches: 137
 Percent Similarity: 33.50% Conservative: 60

523 LysGluThrAlaProThrThrProIysIleuThrProThrMet-
||| ||| |||:::

COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/757,022B
 FILING DATE: 19910910
 CLASSIFICATION: 530
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cserr, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 50:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1314 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-50

Alignment Scores:
 Pred. No.: 7.01e-12 Length: 1314
 Score: 285.50 Matches: 137
 Percent Similarity: 33.50% Conservative: 60
 Best Local Similarity: 23.30% Mismatches: 240
 Query Match: 8.65% Indels: 151
 DB: Gaps: 24

US-10-009-782-1 (1-1758) x US-07-757-022B-50 (1-1314)

```

QY 32 CCATGTCACCAATCCGATGCCAGCCCTTCGACCTGCTCGCGGGCGACCCCTCATCG 91
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 ProthrProthrThrlThlYsserAlaProthrThrlProlysgluProAlaProthr 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 92 ACGGACGA-----ACACCCGGGGGGGGCGGCGGACCTGGCGCGCGGCGGAC 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 ThrYsserAlaProthrThrlProlysgluProAlaProthr-----Thr 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 GCATGCCCGCATCGGCGATGTCGACGCCGCGCGACACCGGGGTCGACGTGTGCG 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 ThrYsgluProAlaPro-----ThrlProlysgluProAlaProthr----- 325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 203 GCCGTGTGTGGCGCGGCTTCATCGACTGCGACACCCAGACGACGACCTGCTCA 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 326 -----ThrlYsgluProAlaProthrThrlYsser 337
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 263 GCGCTCGGACATGACGCCAAGATCTGCAAGGGGTCACACCGGTGTGACGGGCAT 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 AlaProthrThrl-----ProlysgluProAlaPro----- 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 323 GCGGCATCAGCGTGGCGCGCTGGCGGACGCCAACCCCGCGCCCTGGACCTGTG 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 348 -----ThrlProlysgluProAlaProthrThrlProlys 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 383 ACGAAGGCGGCTTACGCTTTCAGAGCGCTTCGCGACTACTGACCGCTTGGCGGCA 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 GluProAlaProthrThrlProlysgluProthrProthrThrlProlysgluProAlaPro 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 443 CGCGGGGG-----CCGTCAACGCCGCTGTATGTGGGCATTCACAGCTGCGCG 493
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 380 ThrThrlYsgluProAlaProthrThrlPro----- 389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 494 CCGGGTCATGCGGACTTGCAGCGCGCCCGCACCGAGAAATCCGCGCATGCGGG 553
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 -----LysgluProAlaProthrThrlProlysgluProAlaPro 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 554 ACCGCGCGAGGAGAACCATGGCGGGCGGCATCCGCAATTGGACCGCGCTTAC 613
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 ThrThrlProlysgluProAlaProthrThrlProlys-----gluProAlaProthrThrl 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 614 CGCCCGCGCGCGCGGCA-----CCACCGAAGATCATCGAGTGTGCGCGCGCTGAGCG 670
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ThrYsgluProserProthrThrlProlysgluProAlaProthrThrlYsserAla 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 671 CGCATGGCGGCTATCGACCCACCAATGCGGACGAGGCGGACGACATCGTGGCGCG 730
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 ProthrThrlThrlYsgluProAlaProthrThrlThrlYsserAlaProthrThrlProlys 460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 731 TGGAGGAACCTCCGCGATGCGCGGAGAGCTGAGCGCGGTGATCTGCACACA 790
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 GluProserProthr-----ThrlYsgluProAlaProthrThrl 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 791 AGGTCAATGGCGACCCCAATTCGCGCGCTGCGGAGACGTCGCGTGAATGAGGCGG 850
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 ProlysgluProAlaPro-----ThrlPro 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 851 CCATGCGCGCGCGGACGATCTGCTGAGCGGATTCCTACGTGCGCGGCTCCACCATCG 910
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 LysLysProAlaProthrThrlProlysgluProAlaProthrThrlProlysgluProAla 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 911 TCAGCAGGACCGCGTGTGCTGCGGAGGACGACCATCATACCTGTGTGACAGCCTTCC 970
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 ProthrThrlThrlYsgluProAlaProthrThrlAlaProthrThrlProlysgluProAlaProthrThrlPro 523
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 971 CCGAATGACGCGCGCGGCGGACCTGATGAATGCGCGCGGCGGCAATCCAACTAG 1030
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 524 LysgluThrlAlaProthrThrlProlysgluThrlProthrThrlProgluLysleuAla 543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1031 ACGTGTGCGCGGACGTCGCGGCGCGCGGCGGCGGCTGCTCATATGACGACCGGACG 1090
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 ProthrThrlPro-----gluLysProAlaProthrThrlProgluLysleuAlaProthr 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1091 TGCAGGCAATCTGCGGTGCGCGCGGACCATGATGCG-----GTCGCGAGCGGCGCGGACG 1147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 562 ThrProgluLysProthrThrlProthrThrlProgluLysleuAlaProthrThrl----- 578
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1148 ACGAGCGCGCGCATCGCGCTGTGGGCGACCTCCGCGGCTGTGGGCGCACTATCGCG 1207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 ----- 578
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1208 GCGACTGCGGCTGTCCCGCTGAGAGGCGCGTATGAAATGACCGGCTGACCGCG 1267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 579 -----Pro 579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1268 CGCGCTGCGCTGCGCGGCGCGGCGGACGTGACGCGGCGGACTGCGGACCTGTG 1327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 LysAlaAlaAla-----ProAsnThrlProlysgluProAlaProthrThrlProlysgluPro 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1328 TGTTCAGCCCGCGCGGCTGCGGATACCG-----CCACCTGGAACACCTTACGAGCGG 1384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 AlaProthrThrlProlysgluProAlaProthrThrlProlysgluThrlAlaProthrThrl 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1385 CGCGCGGATCATCTCGGTATGACGCAACGCGCGCGCGGTGTGCGAAGAGCGCTTCA 1444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 ProlysgluThrlAlaPro-----ThrlLeuLysgluProAlaProthrThrlProlys 637
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1445 CGCGCAGACATGCGCGCGCTGCTGCGACGACGCGCGGCTGAGCGCGGCGGCGG 1504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 638 ProAla-----ProlysgluLeuAlaProthrThrlThrlYsgluProthrThrl----- 654
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1505 TACAATCGGCGGTGAAGCGGCGCGGCGCGGCTGCGCGCTCCACACCTGAGCAACCGGCT 1564
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 655 -----ThSerAspLysProAlaProThrThrProLysGlyThrAlaPro 669
QY 1565 ACATGGCCCCCTCCCTCCGCTCGCAATAGCGCCCGCATATCGTGGCAAGAGTGA 1624
    ||| ||| |||
Db 670 ThrThrProLysGluProAlaProThrThrPro-----680
QY 1625 TGGGGCGCGGCTCGCGCGAGCGCAAGCGCCGAAATGACCTGCAAGACCTGTCCG 1684
    ::::: ||| ||| |||
Db 681 -----LysGluProAlaProThrThrProLysGlyThrAlaPro---ThrThrLeuLys 697
    ::::: ||| ||| |||
QY 1685 AGGCCAGCG---GCATCGCGGTCTGACCTGTCTCAAG-----CCGAGCTGGGGC 1732
    ||||| ::::: ||| ||| |||
Db 698 GluProAlaProThrThrProLysProAlaProLysGluLeuAlaProThrThrThr 717
    ::::: ||| ||| |||
QY 1733 AGATCGCCCTGAGCTAGAGAAGC 1756
    ::::: ||| |||
Db 718 LysGlyProThrSerThrThrSer 725

```

Search completed: May 11, 2003, 12:17:20
 Job time : 65.5 secs

and is derived by analysis of the total score distribution.

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution

Result No.	Score	Query Match	Length	DB	ID	Description
1	2511	76.1	484	22	AAB48945	Alcaligenes xylosoxidans
2	2481	75.2	484	21	AAAY84943	Amino acid sequencer
3	942	28.6	526	22	AAB96073	putative P. adyarsii
4	612	18.6	558	22	AAB47456	D-ethanolamine dehydrogenase
5	411	12.5	1212	20	AAH87503	Human N-methyl-D-arginine
6	402.5	12.5	1261	20	AAH87503	Human N-methyl-D-arginine
7	378.5	11.5	1332	21	AAH82639	Human N-methyl-D-arginine
8	378.5	11.5	1332	21	AAH82639	Human N-methyl-D-arginine
9	370	11.2	1081	21	AAH82639	Human NMDA receptor
10	370	11.2	1081	21	AAH82639	Human N-methyl-D-arginine
11	341	10.5	900	22	AAB55523	Human NMDA receptor
12	339	10.3	1592	22	AAB55523	Human NMDA receptor
13	332.5	10.1	999	22	AAB55827	Human NMDA receptor
14	331.5	10.0	1615	22	AAB55827	Human NMDA receptor
15	321.5	9.7	599	22	AAB55827	Human NMDA receptor
16	318	9.8	506	22	AAAY04841	Novel human dihydro
17	318	9.8	1592	22	AAB55827	Novel human dihydro
18	313.5	9.7	599	22	AAB55827	Novel human dihydro
19	312	9.5	558	20	AAAY28184	Novel human dihydro
20	309	9.4	900	22	AAB55523	Novel human dihydro
21	307.5	9.4	999	22	AAB55523	Novel human dihydro
22	307	9.3	1017	22	AAB55827	Novel human dihydro
23	306.5	9.3	572	20	AAAY04954	Novel human dihydro
24	302	9.2	627	22	AAU51580	Novel human dihydro
25	295	9.1	1615	22	AAB55827	Novel human dihydro
26	294	9.1	1061	20	AAH87503	Novel human dihydro
27	290.5	8.8	1098	22	AAB55827	Novel human dihydro
28	289.5	8.8	5179	22	AAH87503	Novel human dihydro
29	287.5	8.9	1017	22	AAH87503	Novel human dihydro
30	286	8.8	1081	21	AAH87503	Novel human dihydro
31	285	8.8	1081	21	AAH87503	Novel human dihydro
32	285.5	8.8	1017	22	AAB55827	Novel human dihydro
33	285.5	8.7	1299	22	AAH87503	Novel human dihydro
34	285.5	8.7	1404	13	AAH87503	Novel human dihydro
35	285.5	8.7	1404	13	AAH87503	Novel human dihydro
36	285.5	8.7	1404	13	AAH87503	Novel human dihydro
37	285.5	8.7	1404	13	AAH87503	Novel human dihydro
38	284.5	8.6	476	22	AAB55827	Novel human dihydro
39	283	8.6	440	22	AAB55827	Novel human dihydro
40	283	8.6	819	22	AAB55827	Novel human dihydro
41	283	8.6	1065	14	AAH87503	Novel human dihydro
42	283	8.6	1065	14	AAH87503	Novel human dihydro
43	283	8.6	1065	14	AAH87503	Novel human dihydro
44	277	8.4	1605	19	AAH87503	Novel human dihydro
45	276	8.4	1040	22	AAB55827	Novel human dihydro

RESULT 1	
ID	AA048975
XX	AA048975 standard; Protein; 484 AA.
AC	AA048975;
DT	27-MAR-2001 (first entry)
XX	
DE	Alcaligenes xylosoxidans subspecies xylosoxidans D-aminocyclase.
KW	D-aminocyclase; zinc tolerant host; recombinant production;
KW	zinc-enhanced expression; D-form amino acid synthesis;
KW	antibiotic production; peptide drug; pharmaceutical manufacturing
XX	
XX	Alcaligenes xylosoxidans.
OS	
PN	MO200078926-A1.
XX	
28-DEC-2000.	

XX 15-JUN-2000; 2000MO-JP03932.
 XX 17-JUN-1999; 99JP-0170555.
 XX (AMANO) AMANO ENZYME INC.
 XX Takeuchi K, Koide Y, Hirose Y, Moriyuchi M, Isobe K;
 DR WPI; 2001-080828/09.
 DR N-PSDB; AAC91797.

XX Transformed microorganism from zinc-tolerant host for selective
 PT production of D-aminoacylase, useful in synthesis of high
 PT optical-purity D-amino-acids for antibiotic side-chains and peptide
 PT drugs.

XX Claim 2; Page 14-18; 22pp; Japanese.

CC The invention relates to a recombinant zinc-tolerant microorganism which
 CC expresses the D-aminoacylase from Alcaligenes xylosoxidans subspecies
 CC xylosoxidans. The presence of zinc ions in the culture medium enhances
 CC expression of the D-aminoacylase gene, and the invention also relates to
 CC the process of recombinantly producing the D-aminoacylase using the
 CC microorganism of the invention. The recombinant microorganism is used
 CC for the selective production of D-aminoacylase, which is useful in the
 CC synthesis of high optical-purity D-form amino acids for use in the
 CC production of antibiotics and peptide drugs. The present sequence
 CC represents Alcaligenes xylosoxidans subspecies xylosoxidans
 CC D-aminoacylase.

XX Sequence 484 AA;

Alignment Scores:

Pred. No.: 4.43e-173 Length: 484
 Score: 2511.00 Matches: 484
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.11% Indels: 0
 DB: 22 Gaps: 0

US-10-009-782-1 (1-1758) x AAB48975 (1-484)

QY 34 ATGTCCCAATCCGATTCCTCCAGCCCTTGACCTGCTGCTCCGGGCGGCACCTTCATCGAC 93
 DB 1 Metserginseraspserglnprophaspseuduleuilaiglylthrleuilaasp 20
 QY 94 GGCAGCAACACCCCGGGGCGCGCGCGACCTGGCGTGGCGCGCGCGACCGACCGCGCC 153
 DB 21 Glyserasnhrproglarargalaaspleuglyvalarglyasphrgillealaala 40
 QY 154 ATGGCGATCTGTGCGAGCGCGCGCGCGACACCGCGGTGCGACGTGTGGGCGCTGTGTC 213
 DB 41 Illeglyaspseuseraspalaalaahisthrarvalasparvalserglyleuval 60
 QY 214 GCGCGCGGCTTCATCGACCTCGACACCGACGACGACACTACCTGCTCAGGGCTCGCGAC 273
 DB 61 AlaProgllypheiilaaspserrhsthrhisaspasparutylleuulargargasp 80
 QY 274 ATAGGCGCAAGATCTCGAGGGCGTCCACACGGTGTGTCAGGCGCAATTCGGCATCGAC 333
 DB 81 Metthrprolyslleserlmglyvalthrthrvalvalthrglyasnsglyllester 100
 QY 334 CTGGCGCGCTGGCGACCGCAACCGCGCGCGCGCGCGACCTGCTGCGACGAGCGCGC 393
 DB 101 LeuAlaProleuAlaHisAlaasnProProAlaProleuAspLeuAspLuglyly 120
 QY 394 TCTTACGCTTTGAGCGCTTCGCGACACTGACGCGCTTGGCGCGCGCGCGCGCGC 453
 DB 121 SerTyrArgPhegluarPheAlaAspTyrleuAspAlaLeuAlaThrProAlaAla 140
 QY 454 GTCAAGCGCGCTGTATGTGGCGCATTCACGCTCGCGCGCGGTGATCGCGACTTG 513

DB 141 ValAsnAlaAlaCysmetValIglyHisSerThrLeuArgAlaAlaValMetProAspLeu 160
 QY 514 CAGCGCGCGCGCGCACGACGAGGAAATCGCGGCATGCGGACCTGGCGCGAGGACGATG 573
 DB 161 GlnArgAlaAlaThrAspLuglylValAlaAlaMetArgAspLeuAlaLuglylMet 180
 QY 574 GCGAGCGCGCATGGCGCTTCGACCGCGCGCGCTTCACCGCGCGCGCGCGCGCGC 633
 DB 181 AlaSerGlyAlaIleIleIleIleSerThrGlyAlaPheTyrProProAlaAlaThr 200
 QY 634 ACCGAGAGATCATGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 693
 DB 201 ThrGluGluIleIleGlyValCysArgProleuSerAlaHisGlyIleTyrAlaThr 220
 QY 694 CACATGGCGCGAGAGCGGAGCGACACTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGC 753
 DB 221 HisMetArgAspGluGlyIleHisIleValAlaAlaLeuGluGluThrPheArgIle 240
 QY 754 CGCGAGCTGAGCTGCCGCTGTGATCTCGACACGACGACGACGACGACGACGACGACG 813
 DB 241 ArgGluLeuAspValProValValIleSerHisIleValMetGlyGlnProAspPhe 260
 QY 814 GCGCGCTGCGCGAGAGCGCTCCGCTGATCGAGCGCGCGCGCGCGCGCGCGCGCGC 873
 DB 261 GlyTyrSerArgGluThrLeuProleuIleGluAlaAlaMetAlaArgGlnAspValSer 280
 QY 874 CTGAGCGCTATCCCTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 933
 DB 281 LeuAspAlaTyrProTyrValAlaIleGlySerThrMetLeuArgGlnAspArgValLeu 300
 QY 934 GCGCGAGCGCACATCATCATCCGCTGTGCGACCCCTCCCGACGACGAGCGCGCGCGC 993
 DB 301 AlaGlyArgThrIleIleIleThrTyrCysLysProPheProIleuSerGlyArgAspLeu 320
 QY 994 GATGAGTCCGC 1053
 DB 321 AspGluValAlaAlaGluArgGlyLysSerLysTyrAspValProGluLeuGlnPro 340
 QY 1054 GCGCGCGCATCTACTCTCATGATGAGGACGACCGCGCGCGCGCGCGCGCGCGCGC 1113
 DB 341 AlaGlyAlaIleTyrPheMetLeuAspGluProAspValGlnThrGlyLeuAlaPheGly 360
 QY 1114 CGAGCATATATGGCTTCGCGACGCGCGCTGCGCGACGACGACGCGCGCGCGCGCGCGC 1173
 DB 361 ProThrMetIleGlySerAspGlyLeuProHisAspGluArgProHisProArgLeuTrp 380
 QY 1174 GGCACCTTCGC 1233
 DB 381 GlyThrPheProArgValLeuGlyHisTyrAlaArgAspLeuGlyLeuPheProLeuGlu 400
 QY 1234 ACGGCGGTATGGAATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1293
 DB 401 ThrAlaValAlaIleTyrPheMetThrGlyLeuThrAlaAlaArgPheGlyLeuAlaGly 420
 QY 1294 CAGCTGACGCGCGCGCTACTTGCGCGACCTGGTGGTGTTCGACCGCGCGCGCGCGAT 1353
 DB 421 GlnLeuGlnAlaGlyTyrPheAlaAspLeuValAlaPheAspProAlaThrValAlaAsp 440
 QY 1354 ACCGCGACTTCGAAACACTACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1413
 DB 441 ThrAlaThrPheGluHisProThrGluArgAlaAlaGlyIleHisSerValTyrValAsn 460
 QY 1414 GCGCGCGCGCTGCGCAAGACGACGCGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGC 1473
 DB 461 GlyAlaProValAlaTrpGlnGluGlnAlaPheThrGlyGlnHisAlaGlyArgValLeuAla 480
 QY 1474 CGCAGCGCGCGC 1485
 DB 481 ArgThrAlaAla 484

RESULT 2
 ID AAB48943 standard; Protein; 484 AA.

XX AC AAY84943;
 XX 21-AUG-2000 (first entry)
 XX DE Amino acid sequence of an aminocyclase enzyme.
 XX DE Aminocyclase; N-acetyl-D-tryptophan; R-N-acetyl-2-thienylalanine;
 XX DE R-N-acetyl-4-chlorophenylalanine; D-amino acid; N-acyl amino acid;
 XX DE pesticide; antibiotic.
 XX OS Alcaligenes sp.
 XX PN WO200023598-A1.
 XX PD 27-APR-2000.
 XX PF 20-OCT-1999; 99WO-GB03458.
 XX PR 20-OCT-1998; 98GB-0022947.
 XX PR 01-APR-1999; 99GB-0007739.
 XX PA (CHIR-) CHIROTECH TECHNOLOGY LTD.
 XX PI Taylor SC, Brown RC;
 XX DR WPI; 2000-339697/29.
 XX DR N-PSDB; AAAL5140.
 XX PS
 XX PT New isolated D-amino acylase enzyme, useful for the preparation of
 PT D-amino acids for use as intermediates in the preparation of
 PT pesticides, antibiotics and other pharmaceuticals
 XX PT
 XX PS
 XX PS Claim 2; Page 20-22; 26pp; English.
 CC The present sequence represents an aminocyclase enzyme. The enzyme is
 CC capable of hydrolysing N-acetyl-D-tryptophan at a substrate
 CC concentration of 10 grams per litre, and exhibiting faster conversion
 CC of R-N-acetyl-2-thienylalanine than of
 CC R-N-acetyl-4-chlorophenylalanine. The enzyme can be used for the
 CC preparation of a D-amino acid by conversion of a corresponding
 CC D-N-acylamino acid. It can also be used for resolving a racemic
 CC mixture of N-acylamino acids and deprotecting optically-enriched
 CC N-acyl amino acids. The D-amino acids are useful as intermediates in
 CC the production of various pesticides, antibiotics and other
 CC pharmaceuticals.
 CC XX
 SQ Sequence 484 AA;
 Alignment Scores:
 Pred. No.: 6,49e-171 Length: 484
 Score: 2481.00 Matches: 478
 Percent Similarity: 99.388 Conservative: 3
 Best Local Similarity: 98.768 Mismatches: 3
 Query Match: 75.20% Indels: 0
 DB: 21 Gaps: 0
 US-10-009-782-1 (1-1758) x AAY84943 (1-484)
 QY 34 ATGTCCCAATCCGATTCGCCAGCCCTGCTGCTGCGGGGCGACCCCTCATGCAC 93
 DB 1 MetAlaGlnSerAspSerGlnProPheAspLeuLeuValGlyGlyThrLeuIleAsp 20
 QY 94 GGCAGCAACACCCCGGGGGGGGGCGGCGGACCTGGGCTGGCGGCGACCGCATGCGCC 153
 DB 21 GlySerAsnThrProGlyAlaGlyAlaAspLeuGlyValArgGlyAspArgIleAlaAla 40
 QY 154 ATGGCGATCTGTGGAGCGCGCGCGCACACCGGCTGACGTCGTCGTCGTCGTCGTC 213
 DB 41 IleGlyAspLeuSerAspAlaAlaAlaAlaHisThrArgValAspValSerGlyLeuValVal 60
 QY 214 GCGCGCGGCTCATCTGCACTGGCACCGCCAGCAGACAACTACTGCTCAGCGCTCGCGAC 273

DB 61 AlaProGlyPheIleAspSerHisThrHisAspAspAsnTyrLeuLeuArgArgAsp 80
 QY 274 ATGACGCCCAAGATCTCGCAGAGCGCTCACACGCTGCTCAGCGGCAATTGCGCATACG 333
 DB 81 MetThrProLysIleSerIndGlyValThrThrValValThrGlyAsnGlyIleSer 100
 QY 334 CTGGCGCGCTGGCGCGCACCGCAACCGCGCGCGCGCGCGCGCTGCTGACGAAAGCGCG 393
 DB 101 LeuAlaProLeuAlaHisAlaAsnProPheAlaProLeuAspLeuAspIleGly 120
 QY 394 TCTTACCGTTTGAGCGCTTCCCGCATACCTGACAGCGCTTGGCGCACCGCGCGCG 453
 DB 121 SerTyrArgPheGlyArgPheAlaAspTyrLeuAspAlaLeuArgAlaThrProAla 140
 QY 454 GTCAACGCGCGCTGATGGTGGCGCATTCACGCTGCGCGCGCGCTCATGCGGACTTG 513
 DB 141 ValAsnAlaIleCysMetValGlyHisSerThrLeuAlaGlyAlaValMetProAspLeu 160
 QY 514 CAGCGCGCGCGCGCGCACCGAGCAATCGCGCGCATGCGGACCTGGCGGAGGAGCGCATG 573
 DB 161 GlnArgAlaAlaThrAspGluGluIleAlaAlaMetArgAspLeuAlaGluAlaMet 180
 QY 574 GCCAGCGCGCGCATGCGCATTTCCAGCGCGCGCTTACCGCGCGCGCGCGCGCGCAC 633
 DB 181 AlaSerGlyAlaIleGlyIleSerThrGlyAlaPheTyrProPheAlaAlaArgAlaThr 200
 QY 634 ACCGAAAGATCATGAGAGTGGCGCGCGCGCGCGCGCATGCGCGCATGCGCGCATGCGCAC 693
 DB 201 ThrGluGluIleIleGlyValCysArgProLeuSerAlaHisGlyGlyIleTyrAlaThr 220
 QY 694 CACATGCGCGCAGAGAGCGGACACATCGTGGCGCGCGCTGAGAGAACTTCGCGCATGCG 753
 DB 221 HisMetArgAspGluArgGluGluHisIleValAlaAlaLeuGluGluThrPheArgIleArg 240
 QY 754 GCGGAGCTGGACGTCGCGCGCGCTGATCTCGCACCAAGTTCATGCGCGCAATTC 813
 DB 241 ArgLysLeuAspValProValIleSerHisLysValMetGlyGlnProAspPhe 260
 QY 814 GCGCGCTGCGCGGACGCTCGCTGATGACGCGCGCGCATGCGCGCATGCGCGCATGCG 873
 DB 261 GlyArgSerArgGluThrLeuProLeuIleGluAlaAlaMetAlaArgGlnAspValSer 280
 QY 874 CTGACGCGGATTCCTACGTCGCGCGCGCTCCACATGCTCAAGCAGGACCGCTGCTGCTG 933
 DB 281 LeuAspAlaTyrProTyrValAlaGlySerThrMetLeuLysGlnAspArgValLeuLeu 300
 QY 934 GCGGAGCGCAATCATCACTGCTGCAAGCGCTTCCCGCACTGACGGCGCGCATG 993
 DB 301 AlaGlyArgThrIleIleThrTyrCysValArgProPheProLeuLeuSerGlyArgAspLeu 320
 QY 994 GATGAAGTTCGCGCGCGCGCGCAATTCGAAGTTCAGTGGCGCGCGCGCGCGCGCG 1053
 DB 321 AspGluValAlaAlaGluArgGlyLysSerLysTyrAspArgValAlaProGluLeuGlnPro 340
 QY 1054 GCGGCGCGCATCTACTCATGATGAGAGCAACCGGACGTGAGCGCATCTGCGCTGCG 1113
 DB 341 AlaGlyAlaIleTyrPheMetMetAspGluProAspArgAlaGlnArgIleLeuAlaPheGly 360
 QY 1114 CCGACATGATGGCTCCG 1173
 DB 361 ProThrMetIleGlySerAspGlyLeuProHisAspGluArgProHisProArgLeuTyr 380
 QY 1174 GGCACCTTCCCGCGGCTGTGGGGCACTATGCGCGGACCTGGGCTGTCGCGCGGAG 1233
 DB 381 GlyThrPheProArgValLeuGlyHisTyrAlaArgAspLeuGlyLeuPheProLeuGly 400
 QY 1234 ACGCGGATGGAAGATGACCGGCTGACCGCGCGCGCGCTTGGCGCGCGCGCGCG 1293
 DB 401 ThrAlaValTyrLysMetThrGlyLeuThrAlaAlaArgPheGlyLeuAlaGlyArgGly 420
 QY 1294 CAGCTCAGCGCGGTAATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1353
 DB 421 GlnLeuGlnAlaGlyTyrPheAlaAspLeuValValPheAspProAlaThrValAlaAsp 440

QY 1048 -----CACCCGCGCGCCGACATCTACTCATGATGAGAACCGACGCG 1092
 Db 376 LeuAlaLysAspGlyThrAsnAlaGlyMetLeuValPheLeuMetSerGluGlnAspVal 395
 QY 1093 CAGCGCATCTCGGCGTGGCCGACCATGATCGGCTCCGAGCGGCTCCGACGACGAG 1152
 Db 396 GluArgIleuSerHisProTyrSerMetIleGlyThrAspGlyLeuAspSerGlyGlu 415
 QY 1153 ---CGCCGCGATCGCGCGCTGGGGGACCTTCCCGCGGCTGCTGGGCGACTATCGCGC 1209
 Db 416 GlyLeuProHisProAlaGlyAlaTyrGlyThrPheProArgValIleuGlyArgTyrValArg 435
 QY 1210 GACCTGGGCGCTGTTCGCGCTGAGACGCGGCTATGAGAGATGACGCGGCTGACGCGCGC 1269
 Db 436 GluTyrLysLeuLeuArgLeuGlnAspAlaIleArgLysMetThrSerLeuProAlaLeu 455
 QY 1270 CGCTTCGCGCTGGCGCGCGGCGGACGCTGACGCGCGGCTACTTCGCGGACTGGTGTG 1329
 Db 456 LysLeuGlyLeuLysAspArgGlyLeuValLysGluGlyMetThrPalaAspLeuValIle 475
 QY 1330 TTCGACCGCGCGACGCTGGCGGACATCGCGCTTCGACACACCTTACGACGCGCGCGC 1389
 Db 476 PheAspProHisArgValLysAspArgAlaThrTyrThrAsnProArgLeuProProasp 495
 QY 1390 GGCATCCATTCGCTGATGATCAACGCGCGCGGCTGTGCAAGAGACGCGCTTACGCGCGC 1449
 Db 496 GlyIleuTyrValIleValAsnGlyValLeuSerValGluAsnGlyGluLeuThrGly 515
 QY 1450 CACGATGCGCGCGCGGCTGCTGCGACGCGCGC 1482
 Db 516 AspAlaGlyValValIleArgArgThrSer 526
 RESULT 4
 AAB47456
 ID AAB47456 standard; Protein; 558 AA.
 XX AAB47456;
 AC
 XX 13-DEC-2001 (first entry)
 DT
 XX D-aminoacylase.
 DE
 XX D-aminoacylase: N-acetyl-D-tryptophan; N-acetyl-D-phenylalanine; PCR;
 KM N-acetyl-D-valine; N-acetyl-D-leucine; N-acetyl-D-methionine; primer;
 KW probe; synthesis; detection; D-amino acid; substrate specific; amplify;
 thermal stability; polymerase chain reaction.
 OS Hypomyces mycophilus.
 XX
 OS
 PN EP1120465-A1.
 PD
 XX 01-AUG-2001.
 PD
 XX 25-JAN-2001; 2001EP-0101739.
 PF
 XX 27-JAN-2000; 2000JP-0019080.
 PR 22-MAY-2000; 2000JP-0150578.
 XX
 XX (DAIL) DAICEL CHEM IND LTD.
 PA
 XX Mitsuhashi K, Yamamoto H, Matsuyama A, Tokuyama S;
 PI
 XX WPI; 2001-551332/62.
 DR N-PSDB; AAH43261.
 XX
 PT Novel D-aminoacylase-encoding gene derived from filamentous fungus
 PT Hypomyces mycophilus, useful for producing D-tryptophan from
 PT N-acetyl-D-tryptophan, useful as medicinal raw material
 XX
 PS Claim 1; Page 17-20; 33pp; English.
 CC This sequence shows a D-aminoacylase polypeptide. D-aminoacylase has

CC physicochemical properties that include the action of the enzyme on
 CC N-acetyl-D-amino acids to produce the corresponding D-amino acids,
 CC and substrate specificity, where the enzyme acts on
 CC N-acetyl-D-tryptophan, N-acetyl-D-phenylalanine, N-acetyl-D-valine,
 CC N-acetyl-D-leucine, and N-acetyl-D-methionine, but not on
 CC N-acetyl-L-tryptophan, N-acetyl-L-phenylalanine, N-acetyl-L-valine,
 CC N-acetyl-L-leucine, or N-acetyl-L-methionine. Fragments of the
 CC D-aminoacylase cDNA are useful as primers or as probes for synthesizing
 CC or detecting the full length cDNA. D-aminoacylase is useful for
 CC producing D-amino acids. The enzyme has substrate specificity,
 CC thermal stability and produces D-amino acid efficiently by incubating
 CC the fungus derived D-aminoacylase with N-acetyl-D-amino acid under
 CC proper conditions. The enzyme has high enzymatic activity for
 CC N-acetyl-D-tryptophan and is excellent in industrial applicability.
 CC The recombinant polypeptide of D-aminoacylase can be manufactured at
 CC a low cost and in large quantities.

Sequence 558 AA;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
61	612.00	558	171	76	226	13	
4	45.328	558	76	226	13		
Best Local Similarity:	31.388						
Query Match:	18.538						
DB:	22						

US-10-009-782-1 (1-1758) x AAB47456 (1-558)

QY 61 GACCTGCTGCTCGCGCGCGGACCCATGATGACGACGACGCGCGCGGCGG---CGC 117
 Db 4 GluIleuPheHisSerAlaThrValIleThrGlyAspIleAlaGlnProPheVal 23
 QY 118 GCGGACCTGGCGCGCGCGCGGACCGGACCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 171
 Db 24 AlaAspValLeuValSerLysGlyLeuIleAlaLysIleGlyAsnProGlySerIleAsn 43
 QY 172 GCGCGCGCGGACCGCGG---GTGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 228
 Db 44 AlaThrProAspThrArgHisLeuAspValThrGlyTyrIleLeuSerProGlyPheIle 63
 QY 229 GACTCGGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 288
 Db 64 AspMetHisAlaHisSerAspLeuTyrLeuLeuSerHisProAspHisGluAlaLysIle 83
 QY 289 TCGGAGGCGGCTCAACGAGTGTACGCGGCAATTGCGGATGAGCTGGCGGCGGCGGCGG 348
 Db 84 ThrGlnGlyCysThrThrGluValValGlyGlnAspGlyIleSerTyrAlaProIleArg 103
 QY 349 CAGCGCACCGG-----CGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 372
 Db 104 AsnValAspIleLeuArgAlaIleArgGlnIleAlaGlyTyrAsnGlyAsnProThr 123
 QY 373 GAC-----CTGCTGGACGAGAGCGGCTCTTACCGTTTTCGAGCGCTTCCGCG 417
 Db 124 AspGluGluCysArgThrThrLeuLysGlyValGlyMetPheGluThrIleThrIleGly 143
 QY 418 GACTACTGAGCGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477
 Db 144 GluTyrLeuAspCysLeuGluArgAsnArgThrAlaThrAsnValAlaMetLeuValPro 163
 QY 478 CATTCACGCTCGCGCGCGGCGGCTCATGCGGACTTGCAGCGCGGCGGCGGCGGCGGAA 537
 Db 164 GlnGlyAsnLeuArgLeuLeuAlaCysGlyProTyrAspThrProAlaSerAlaGluIle 183
 QY 538 ATCGGCGCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
 Db 184 IleGlnAspGlnIleGlnLeuLeuArgGluAlaMetAlaGlnGlyValGlyMetSer 203
 QY 598 ACCGCGCGCTTACCG 657
 Db 204 SerGlyLeuThrTyrThrProGlyMetTyrAlaSerThrSerGluLeuAlaSerLeuGly 223

```

QY 658 CGGCCGCTGAGCGGCAT-----GGCGCATCTAGCGCCACCATCGCGAGAGAGC 711
    |||
Db 224 AAlaIleuAlaIngluPherProGlyAlaPheThyAlaProHISAlaSerTyrGly 243
QY 712 GAGCAGCATCTGGCCGCGCTGAGAGAACTTCGCGATCGCGCGAGCTGAGCTGGCG 771
    |||
Db 244 PheGlnAlaIleGluSerTyrAlaGluMetLeuAspLeuGlyGluSerThrGlyCysPro 263
QY 772 GTGGGATCTGCGACCAACAGATGAGGCCACGCCCAATTCGGCCGCTCGCGAGAGC 831
    |||
Db 264 ILeuHISLeuThiAlaThiLeuAsnPheserGluAsnGlyGlyAlaProValLeu 283
QY 832 CTGCGCTGATGAGCGCCGCTGAGCGCGC---CAGAGCGATCGCTGAGCGCGCTATCCC 888
    |||
Db 284 ILeuMetValAspLysSerLeuAlaGlyAlaAspValThrLeuAspThrTyrPro 303
QY 889 TACGTGGCGCGCTCCACCATGCTC-----912
    |||
Db 304 TyrLeuProGlyCysThrThrLeuAlaAlaLeuLeuProSerThrAlaSerAlaGlyGly 323
QY 913 -----AAGAGAGAGCGCGCTGCTGCGC 936
    |||
Db 324 ProGlnIuThrLeuLysArgLeuGluAspAlaGluSerArgLysIleArgIleAla 343
QY 937 -----GGAGCGACCATCATCACC---TGTGCAAG-----963
    |||
Db 344 ValGluIleLysGlyCysAspArgLysGlyIleProThrAsnThrAspGluIleGln 363
QY 964 -----CCCTCCCGAAGTGAAGCGCGCGCGAGCTGTGATGAATC 1002
    |||
Db 364 ILeuGlyThrThrasnGluProSerIleAlaSerTyrSerIleArgLysLeuSerGluVal 383
QY 1003 GCGCGCGAGCGCGCAATTCAGTACGAGCTGTGCGCGAGCTG-----1047
    |||
Db 384 AlAGlnSerValGlyLysProThrIleGluValAlaPheGluIleLeuGlnLysAspLys 403
QY 1048 CAGCGCGCGCGCGCATCTTCTATGATGAGAGAAACCGAGCTGAGCGCATCTGCGG 1107
    |||
Db 404 LeuAlaThrSerCysIleMetHISValGlyAsnGluGluAsnValArgIleMetGln 423
QY 1108 TTGCGCGCGCATGATCGCGCTCGAGCGCGCTCGCGAGAGCGCGCATCTGCGCGC 1167
    |||
Db 424 HISArgValHISMetAlaGlySerAspGlyIleLeuHISGlyIleThrLeuHISProArg 443
QY 1168 CTGTGGGAGCACTTCCCGGCGGTCTGCGGCACTATGAGCGCGAGCTGTGCGCTGCCG 1227
    |||
Db 444 AlaTyrGlyThrPheThrArgTyrLeuGlyHISThrSerArgGluLeuSerLeuValAla 463
QY 1228 CTGGAGACGCGCGGTATGAGATGACCGCGCTGACCGCGCGCTGCGCGCTG---GCC 1284
    |||
Db 464 LeuProSerMetIleAlaHISLeuThrSerArgProAlaLysArgLeuSerValTyrPro 483
QY 1285 GGGCGGCGCGCATGCGCGCGGTACTTCGCGCGCACTGTGCTGACCGCGCGCAG 1344
    |||
Db 484 TyrArgGlyLeuIleAlaGluIleGlySerAlaAlaAspIleValAlaPheAsnProGluThr 503
QY 1345 GTGGCGGATACCGCATCTTCGAGACACCTTACGAGCGCGCGCGCGCATCATCTCTG 1404
    |||
Db 504 ValLysAspMetSerThrTyrGluIuProLysValProSerArgGlyIleArgPheVal 523
QY 1405 TACGTCAACGCGCGCGGTCTGCGCAAGAGCGCGTTCACCGCGCGCATGCGCGCGC 1464
    |||
Db 524 LeuValAsnGlyGlnIleAlaValAspGluGlyLysMetThrGlyThrArgGlyGlyLys 543
QY 1465 GTGCTGCGCGCAGC 1479
    |||
Db 544 ThrLeuArgArgSer 548

```

RESULT 5
 AA087503
 ID AA087503 standard; Protein; 1212 AA.
 XX
 AC AA087503;

```

XX 23-FEB-1999 (first entry)
DT Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA22.
DE
XX Human N-methyl-D-aspartate receptor; NMDAR2C;
KW NMDA-activated cation-selective ion channel; glutamate receptor.
XX Homo sapiens.
OS
XX
XX US5849895-A.
PN
XX 15-DEC-1998.
PD
XX 20-APR-1994; 94US-0231193.
PF
XX 20-APR-1994; 94US-0231193.
PR
XX 20-APR-1993; 93US-0052449.
XX
XX (SIBI-) SIBIA NEUROSCIENCES INC.
PA
XX Daggett LP, Lu C;
PI
XX MPI: 1999-069812/06.
DR
XX N-PSDB; AA082909.
XX
XX DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
XX assembly of functional glutamate receptor subunits
XX
XX Example 3; Columns 253-262; 203pp; English.
XX
XX The present sequence represents a human N-methyl-D-aspartate (NMDA)
XX receptor subunit (NMDAR). The nucleic acid sequence does not contain
XX the 366 5' most nucleotides, by the insertion of 11 nucleotides between
XX nucleotides 1300 and 1301, nor the 15 nucleotides at positions
XX 1960-1974, nor the 1061 3' nucleotides, as set forth in AA082889. The
XX cDNA sequence is derived from clone NMDA21. The NMDAR subunits contribute
XX to the formation of NMDA-activated cation-selective ion channels. In
XX addition to being useful for the production of NMDA receptor subunit
XX proteins, the nucleic acids are also useful as probes to identify and
XX isolate nucleic acids encoding related receptor subunits. Functional
XX glutamate receptors can be assembled from several NMDA receptor subunit
XX proteins of one type (homomeric) or from combinations of subunit proteins
XX of different types (heteromeric). The present invention also comprises
XX methods for using such receptor subunits to identify and characterize
XX compounds which affect the function of such receptors, e.g. agonists,
XX antagonists and modulators of glutamate receptor function. The invention
XX also comprises methods for determining whether unknown protein(s) are
XX functional as NMDA receptor subunits.
XX
XX Sequence 1212 AA;
XX
XX
XX Alignment Scores:
XX Pred. No.: 2,396-21 Length: 1212
XX Score: 411.00 Matches: 188
XX Percent Similarity: 33.61% Conservative: 59
XX Best Local Similarity: 25.58% Mismatches: 230
XX Query Match: 12,468 Indels: 258
XX DB: 20 Gaps: 41
XX
XX US-10-009-782-1 (1-1758) x AA087503 (1-1212)
XX
XX QY 32 CCATGTCC-----AATCGATT-----CCGAGC 55
    |||
Db 370 ProCysProAlaAlaGlyArgAlaThrThrProSerAlaAlaGlyThrTrpProThr 389
QY 56 CCTGAGCATGCTGCTGCGCGCGCGCACCTCATGAGCGAGCAGACACCCCG---GCC 112
    |||
Db 390 ProSerSerAlaValArgAspSerAlaSerThrSerArgSerTrpProGluTrpSer 409
QY 113 GCGCGCGCG-----ACCTGGCGCTGCGCGCGCGAGCGAGCGCGCGCATGCGGATC 163
    |||
Db 410 AsnSerProThrThrCysThrThrProThrAlaSerMetAlaSerGlyCysAlaAlaTyr 429

```

QY	164	-----TCTCCAGACCGCCGCGACACACCCGGGTGCAGCTGTCCGGCC	205
Db	430	GlyThrAlaLeuGlyArgCysThrThrSerGlyGln-----	442
QY	206	TGGTGGTCGGCCCGGCTTATGCATCGGCACACCCAGACACAACTACTGCTCAGGC	265
Db	443	TrpProSerAlaProSerProSerMetAlaGln-----	453
QY	266	GTCCGCACATGAGCCCAAGATCTGCAGAGGGGTGCACACAGGTGGTCACGGCAATTGGC	325
Db	454	-----AlaProArgSerThrSerLeuTyrProLeuTyrPargArg-----	466
QY	326	GCATCACCCTGGCGCGCGGTACGGCCAGCCACACCCGCCG-----CCCCCTGGACCTGCTGG	382
Db	467	AlaSerValTyr-----TrpLeuAlaAlaMetAlaProSerProProAlaGlyProSerTyr	484
QY	383	ACGAGAGCGGCTTTCACGTTTCGAGCGCTTCGCCGACTGACGACGCTGGGGGCCA	442
Db	485	SerIstAlaLeu-----GlnCysGlyCys	493
QY	443	CGCGCGCGCGCCCAACGCGCGCTGTATGGAGGGCGCATTCACCTCGCGCGCGGTCA	502
Db	494	Leu-----SerCysAlaSerLeuTyrTrpPro-----SerProSer	506
QY	503	TGCGGACACTTGACAGCGCGCGCCGACCCAGCAGGAATCGCGGCCA-----	547
Db	507	CysSerSerThrSerAlaLeuSerAlaThrThrArgThrSerProGluAlaArgLeuSer	526
QY	548	-----TGCGGACCTCG-----	559
Db	527	LeuSerAlaSerProCysGlyCysSerGlyArgTyrTrpSerSerThrThrGlnCysProSer	546
QY	560	-----CCAGAGAAGCCAAG-----CCAGGGGGCGCATGG	589
Db	547	ArgThrArgGlyAlaProProAlaArgSerTyrPheTyrPheTyrAlaProSerLeuLeuSer	566
QY	590	GCATTCGACCGCGCGCTTACACCGCGCG-----CCGCCCGCGCCACACCGAAGAGA	643
Db	567	Ser-----SerSerProSerPthrArgProThrTyrTrpProProSerSerIysSerAsnThr	584
QY	644	TGATCGAGCTGT-----	655
Db	585	SerThrLeuCysArgAlaSerValThrArgSerPheSerClyLeuLysIleSerThrHis	604
QY	656	-----GCCGCGCGCTGACGGCGCATGGCGGCATCTACG	688
Db	605	LeuSerAlaSerAlaArgCysProThrAlaAlaArgSerGlyThrSerAlaValThrThr	624
QY	689	-----CCACCACATCGCGCGAGCAAGGAGACATCTGGCGCGCGCTGAGGAAA	739
Db	625	ValThrCysThrProThrTrpSerSerSerThrSerAlaArgTyr--ArgThrArgSer	643
QY	740	CGTTCGCGATCGCGCGGACCTGG-----ACGCGCGCG	772
Db	644	ProAlaSerArgTyrGlySerTyrMetProSerSerMetMetLeuLeuSerSerThrThr	663
QY	773	TGCTGATCTCGCACACAAAGTATGGCGACCCCAATTGGCGCGCTCGC-----	823
Db	664	TrpGlnAlaArgThrArgAlaAlaSerTyrPheProLeuGlyLeuAlaArgSerLeuLeu	683
QY	824	-----GGGAGACGTCGCCG-----TGATCGAGCGCGCCATGGCGC	859
Db	684	ProLeuAlaThrAlaSerProCysArgArgThrProThrGlySerGlyProThrTyrArg	703
QY	860	GCCGAGAGCTCT-----	871
Db	704	SerCysSerSerTyrPheGlyThrGluArgHisArgAsnTyrPargGlnCysGlySerGlnGly	723
QY	872	-----CGCTGAGCGCGTATCCCTACGTGCGCGCGCTCCACACATGCTCAAGCAGG	919
Db	724	SerAlaArgMetArgTyrThrArgAlaAlaSerTyrPthrSerThrThrTrpGluAlaSer	743

QY	920	ACGGCGTCTGCTGGCCGGAGCCAGCCATCATCCTGGTGGAAAGCCCTGCCGAATGA	979
Db	744	748
QY	980	GCGGGCGCGCACTGGATGAAAGTGGGGCGGAGACCGCGGAATCCAACTACAGCTGTGC	1035
Db	749	764
QY	1040	CCG---AGTCGACCGCGCGCGGCATCTCATGATGAGCAAGAACCGAGCTGCAC	1099
Db	765	779
QY	1097	GCATCGGGGTGGGGCGCGGACCAATGATCGGCTCCGACGGC---TGCCGACAGCAAC	1153
Db	780	799
QY	1154	GCCCGCATCCGGCGCTGTGGGGCACTTCCCGCGGTGTGGGGCACTATGGCGGCAC	1213
Db	800	815
QY	1214	TGGCGCTGTTCGGGTGGAGAGCGCGGTATGGAAGATGACCGGCGCTGACCGCGCGCT	1273
Db	816	831
QY	1274	TCGGCGCTGGCGGGC-----GGGGCGACGTCC	1300
Db	832	851
QY	1301	AGCGCGGACTGCGCGGACCTGGTGTGTGCACCGCGGCACGGTGGCGCATACCGCA	1360
Db	852	871
QY	1361	CCTTGAACACCCCTACGAGCGCGCGCGCGCATTCATTCGTGTACCTCAACGGCGC	1420
Db	872	891
QY	1421	CGGTGGCAAGACAGCGCTTACCGCGCAGCATCCGCGCGCTGCTGCACGACGG	1480
Db	892	911
QY	1481	CCGCGTGAAGCGCGCGCGCATCTTACATCCGGGGGTAAAGGGCGCGCTGCCGCCCC	1540
Db	912	928
QY	1541	TCCCAACCCCTGGAGCAACCGCTACATGCGCCCTCCCTCGCTGCAATAGCGCCAC	1600
Db	929	948
QY	1600	-----	1600
Db	949	968
QY	1601	1651
Db	969	982
QY	1652	AGG-----	1699
Db	983	999
QY	1700	CGGTCTGCA-----	1735
Db	1000	1013
RESULT 6			
AAW87504			
AAW87504	standard; Proteln; 1061 AA.		
AAW87504;			
23-FEB-1999	(first entry)		
Human N-methyl-D-aspartate receptor subunit encoded by clone INMDA24.			

The present sequence represents a human N-methyl-D-aspartate (NMDA) receptor subunit (NMDAR). The nucleic acid sequence does not contain the 860 5'-most nucleotides, has an additional 11 nucleotides (AAV82891) between nucleotides 1300 and 1301, an additional 24 nucleotides (AAV82890) inserted between nucleotides 23050 and 2351, as set forth in AAV82889. The cDNA sequence is derived from clone NMDA24. The NMDAR subunits contribute to the formation of NMDA-activated cation-selective ion channels. In addition to being useful for the production of NMDA receptor subunit proteins, the nucleic acids are also useful as probes to identify and isolate nucleic acids encoding related receptor subunits. Functional glutamate receptors can be assembled from several NMDA receptor subunit proteins of one type (homomeric) or from combinations of subunit proteins of different types (heteromeric). The present invention also comprises methods for using such receptor subunits to identify and characterise compounds which affect the function of such receptors, e.g., agonists, antagonists and modulators of glutamate receptor function. The invention also comprises methods for determining whether unknown protein(s) are functional as NMDA receptor subunits.

US-10-009-782-1 (1-1758) x AAW87504 (1-1061)

QY	32	CCATGTGCC-----AATCGATT-----CCAGC	55
Db	206	ProCysProAlaAlaGlyAlaThrThrThrProSerAlaAlaGlyThrTrpProThr	222
QY	56	CCCTTGACCTGCTCTCGCGGCGGACCCCTATCGACGGACGACACCCCGG---GGC	112
Db	226	ProSerSerAlaValArgAspSerAlaSerThrSerSerArgSerTrpProGluTrpSer	245
QY	113	GGCGGGCGG-----ACCTGGCGGTGGCGGCGACCGCATGCCCATCGGCGATC	163
Db	246	AsnSerProThrThrCysThrTrpProThrAlaSerMetAlaSerGlyAlaAlaTyr	265
QY	164	-----TGTCGACGCGCGCGCGGACACCCGGGTGCACCTGTGGGCC	205
Db	266	GlyThrAlaLeuGlyArgCysThrThrSerGlyGlnThr-----	278

QY	206	TGTTGTTGTCGGCCGGCCCTTCACATCGACACCCAGACAGACAACCTACTGCTCAGGC	265
Db	279	TrrProSerAlaProSerProSerMetIaYasn	289
QY	266	GTTCGCACATGACGCCAAGATCTGCAGAGGGCGTCCACCACGGTGTACAGGCAATTCG	325
Db	290	-----AlaAlaArgSerThrSerLeuTyrProLeuTyrPAlaArgY-----	302
QY	326	GCATCAGCCTGGCGCCGCTGGCGGACGCCAACCCGCCG--CCCCCTTGACCTGCTGG	382
Db	303	AlaSerValTrrP-----TrrPLeuAlaIaMetAlaProSerProPAlaArgProSerTrrP	320
QY	383	ACGAAAGCGCGCTTACCGCTTTCAGACGCTTCGCCGACTACTGTGACGGCGTGGGGCCA	442
	:::		
Db	321	SerHisIleAlaLeu-----GlnCysGlyCys	329
QY	443	CCGCGGCGGCGCTCAACGCCCGCCCTGTATGTGGGCCATTCACGCTCGCGCGCGGTCA	502
		:::	
Db	330	Leu-----SerCysAlaSerLeuTrrPrrPro-----SerProSer	342
QY	503	TGCCGGACTTCAGCGCGCGCCGACCGCAGGAATCGGGCCA	547
		:::	
Db	343	CysSerSerThrSerAlaLeuSerAlaThrThrArgThrSerProGluAlaArgSerPro	362
QY	548	-----TGGGGACCTGG-----	559
Db	363	GlyAlaGlnLeuSerLeuSerAlaSerProCysGlyCysGlyAlaArgPrrSerThr	382
QY	560	-----CCGAGGAACCATGG-----	574
Db	383	ThrGlnCysProSerAlaGlyThrArgGlyAlaProProAlaArgPrrPrrPrrSerThr	402
QY	575	CCAGCGGCGGCATCGGCATTTCGACCGCGCGCTTACCGCGCG--CCGCGCGG	628
Db	403	ProSerLeuLeuSerSer-----SerSerProAspThrArgProThrPrrProProSer	420
QY	629	CCACACCGAGAGATCATCATGAGTCT-	655
Db	421	SerIysSerAsnThrSerThrLeuCysArgAlaSerValThrArgSerPrrSerGlyLeu	440
QY	656	-----GCCGCGCGCTGAGCGGCG	673
Db	441	LysIleSerThrHisLeuSerAlaSerAlaArgCysProThrAlaIaIaArgSerGlyThr	460
QY	674	ATGGCGGCATTCAGCCACCCACATGCGCGCAGACGACACATCGTGGCGGCGTGG	733
Db	461	SerIaValTrrThrValThrCysThrProThrTrrPrrSerSerThrSerAlaArgPrr	480
QY	734	AGC-----AAACCTTCGCGAATCGGCGCGCGCGCTGG	763
		:::	
Db	481	ArgThrArgSerProAlaSerArgTrrPAlaLeuArgLeuSerLeuSerProGlySerTrrP	500
QY	764	-----ACGTGCGCGGTGGTGAATCTCCGACCAACAAGTCA	796
Db	501	MetProSerSerMetMetLeuLeuSerSerThrTrrPrrGlnAlaArgThrArgAlaIa	520
QY	797	TGGCGCAGCCCATTCGGCGCGCTGGC-----GCCAGACGCTGC	835
Db	521	SerTrrPrrProLeuGlyLeuAlaIaArgSerLeuLeuProLeuAlaThrAlaSerProCys	540
QY	836	CGC-----TGATCGACGCGCGCATGCGCGCGCGACGACGCTT-	871
Db	541	ArgArgThrProThrGlySerGlyProThrTrrPrrArgSerCysSerSerTrrPrrGlu	560
QY	872	-----CGCTGACCGGT	883
Db	561	ArgHisArgAsnTrrPrrGlnCysGlySerGlnGlySerAlaArgMetArgThrArg	580
QY	884	ATCCCTACGTGGCGCGCTCCACCATGCTCAAGCAGACGCGCGCTGCTGGCGCGAGCA	943
	:::	:::	
Db	581	AlaAlaSerTrrPrrSerThrThrTrrPrrGluAlaSerSerThrCysCysTrrP-----	597

```

OY 944 CCATCATCACTGGTGCAGACCCCTCCCGACTGACGGGGCGACCTGGATGAGTGC 1003
DB 598 -----Tripprotrigely 601
OY 1004 CGGCCGAGCGGCCAATCCAGTACAGCTGTGCCG--AGCTGCACCGCGCGCG 1060
DB 602 TTPPCysArgTTPserProglySerThrThrSerThrCysSalatThrArgCys 621
OY 1061 CCATCACTCACTGATGAGACCGACCTGCAGCGCATCCCTGGCCCTGGCCCGCA 1120
DB 622 Pro-----ThrHisProSerThrThrSerThrSerThrLeuSerAlaGlyAla 636
OY 1121 TGCATCGCTCCGACGCGC--TGCCGACACGACGCGCGCCGACCTCCGCTGGGCGCA 1177
DB 637 SerThrAlaAlaSerAlaGlySerAlaSerProAlaHis-----ArgGlyArg 653
OY 1178 CCTCCCGCGGGGTCTGGGGCCTATCGCGGACCTGGCCCTGCTCCGCTGGAGACG 1217
DB 654 ProAlaArgThrSer--ArgProAlaArgProArgProAlaCysSerArgPheCysArg 672
OY 1238 CGGTATGAGAAATGACCGGCTACCGCGCGCTGGCCCTGGCCGCGCGC----- 1288
DB 673 Gln-----ProAlaThrTTPProArgAlaAlaAlaThrProThrAlaPro 688
OY 1289 -----GCGGCGACGCTGCAGCGCGGCTACTGCGCGACCTGG 1324
DB 689 LeuAlaProSerArgTleGlyValAlaAlaAlaValArgProHisArgProProAlaArg 708
OY 1325 TGGTGTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1384
DB 709 ProArgGlyLeuAlaProArgThrAlaHisAlaCysProProProThrArgProGlnSerArgAla 728
OY 1385 CGCGCGCATTCATTCGATCGTACGTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1444
DB 729 ProArgAlaGlyAspArgGlnThrGlyValAlaArgArgLeuGlyAlaGlyLeuArgSer 748
OY 1445 CGGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1504
DB 749 ProArgAlaAlaProArgThrArgGlyArgProCysArgProThrSerProGlyCysArgAla 768
OY 1505 TACATCCGCGCGTGAACGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1564
DB 769 AlaGlnPro-----GlyArgArgGlyGlyArgCysGlyProGlyThrAlaGlyGly 785
OY 1565 ACATGGCCCGCTCCCTCGCGTGCAGATACGCGCGCGCGCGCGCGCGCGCGCGCG 1600
DB 786 ThrSerArgProProSerGlyProCysArgProArgAlaValThrThrAlaProPheLeu 805
OY 1601 -----CCGATATCGTGGCGA 1615
DB 806 GluProThrAspProAlaAlaAlaProSerSerArgSerSerArgSerProArgSerTrp--- 824
OY 1616 AGGAAGTATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1666
DB 825 -----ArgThrCysArgCysSerValArgSerSerThrProGlyGlyArg 839
OY 1667 CCCTGCAAGACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1714
DB 840 Pro-----CysThrArgProGlyProGlyAlaArgAlaArgValThrLeuProCys 856
OY 1715 CCAGGCGCGAGCTGGGCGCA 1735
DB 857 ProAlaPro---TrpProArg 862

```

```

XX Human: N-methyl-D-aspartate receptor; NMDA; NMDAR1A; Ionotropic;
KW glutamate receptor; drug screening; animal model; disease diagnosis;
KW genetic screening.
XX Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 280
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 304
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 326
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 347
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 422
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 438
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 464
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 477
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 501..502
FT /label- unknown
FT /note- "encoded by TGATGA"
FT Misc-difference 590
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 712
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 744..745
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 843
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 848
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 1009
FT /label- unknown
FT /note- "encoded by TAA"
FT Misc-difference 1152
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 1209
FT /label- unknown
FT /note- "encoded by TAG"
FT Misc-difference 1227
FT /label- unknown
FT /note- "encoded by TGA"
FT /label- UNKNOWN
FT /note- "encoded by TGA"
FT /note- "the coding sequence of this protein contains a
number of in-frame stop codons which are represented
by Xaa in the protein sequence"
PN US6111091-A.
XX
XX 29-AUG-2000.
XX
XX 29-SEP-1997; 97US-0940086.
XX
XX 20-APR-1994; 94US-0231193.

```

PR 20-APR-1993; 9305-0052449.
 PA (MERI) MERCK & CO INC.
 PI Daggett LP, Lu C;
 XX
 XX
 DR WPI: 2000-578607/54.
 DR N-PSDB; AAA95032.
 XX

PT Novel DNA fragment encoding human N-methyl-D-aspartate receptor subunit
 PT for identifying mutations and for developing drugs against various
 PT disease states
 XX

Example 3; column 255-264; 205pp; English.

The present sequence is a subunit (designated NMDAR1) of the human
 CC N-methyl-D-aspartate (NMDA) receptor. This is an ionotropic glutamate
 CC receptor which contains cation-specific ligand-gated ion channels. The
 CC receptor and its coding sequence can be used in disease diagnosis and in
 CC research to identify other similar proteins. They can also be used as
 CC probes, for example in genetic screening, and in drug screening, as well
 CC as enabling the production of animal disease models.
 XX

Sequence 1232 AA;

Alignment Scores:
 Pred. NO.: 5,32e-19 Length: 1232
 Score: 378.50 Matches: 189
 Percent Similarity: 33.15% Conservative: 58
 Best Local Similarity: 25.37% Mismatches: 233
 Query Match: 11.47% Indels: 265
 Gaps: 21 44

US-10-009-782-1 (1-1758) x AAB26239 (1-1232)

```

QY 32 CCAATGTCCTC-----AATCCGAT-----CCGAC 55
DB |||||
QY 374 Procyproalalaglyargalathrthproseralalaglythrtrpser 393
DB |||||
QY 56 CATTGACCTGCTGCTGCGCGCGCCGACCTCATCGACGCGACACCCCGG---GGC 112
DB |||||
QY 394 Proseralavalargaspseralaserthrserserargsertrpser 413
DB |||||
QY 113 GCGCGCGC-----ACCTGG-----GCGTGGCGGCG 139
DB |||||
QY 414 AsnserProthrthrcysthrtrp**Prothralasermetlaserlycysalala 433
DB |||||
QY 140 ACCGATCGCCCGCATCGCGCATCTGTCGACGCGCGCGACACCCGCGTGCAGTGT 199
DB |||||
QY 434 Trrglthrthla**Leuglyargcysthrthrserserlyglnthr----- 448
DB |||||
QY 200 CGGCGCTGCTGCTGCGCGCGCGCTCATCGACTCGC---ACACCGACGACGACACTAC 256
DB |||||
QY 449 -----TrrproseralaproserProsermetlarginalaproargser**Thrs 466
DB |||||
QY 257 Tgctcagcgcgtgcgacatgacgacagatctgcagcggtcaccacgctgtgcagcg 316
DB |||||
QY 467 Leu-----TrrProleuthrphargalaserVal**Trr----- 478
DB |||||
QY 317 GCAATGCGGATCAGCCTGCGCGCGCTGCGCGACGCCAACCCGCGG---CCGCCCTG 373
DB |||||
QY 479 -----TrrleuhalmetlaserProserProarg 489
DB |||||
QY 374 ACTGCTGACGAGAGGCGCTTACGCTTTCAGAGCGCTTCGCGACTACCTGACGCGT 433
DB |||||
QY 490 Prosertrpserhalstlealau-----Gln 498
DB |||||
QY 434 TGGCGCCACGCGCGCGCGCTCAACGCGCGCTATGTTGGGACATCAACGCTGCGCG 493
DB |||||
QY 499 Cysgly-----CysleuserCysalaserleuthrtrpPro-----Ser 513
DB |||||
QY 494 CCGCGCTCATCGCGACTTCAGCGCGCGCGCACGACGAGAAATCGCGCGCA----- 547
DB |||||

```

```

DB 514 ProserCysSerSerThrSerAlaLeuSerAlaThrThrArgThrSerProGluAla 533
QY 548 -----TGGCGAGACTCG----- 559
DB 534 ArgLeuSerLeuSerAlaSerProCysGlyCysGlyArgThrSerSerThrThrGln 553
QY 560 -----CCAGAGAACCAATG----- 580
DB 554 CysProSerArgThrArgGlyAlaProProAlaArgSerThrThrSerGlyProSer 573
QY 581 GCGCATCGGATTCGACCGCGCTTACCGCGCG-----CCGCCCGCGCACCA 634
DB |||||
QY 574 LeuLeuSerSer-----SerSerProAlaThrArgProThrTrpProProSer***Ser 591
DB |||||
QY 635 CCGAGACAG----- 643
DB 592 LysSerAsnThrSerThrLeuCysArgAlaSerValThrArgSerPheSerGlyLeuLys 611
QY 644 -----TCATCGAGCTGT-----GCCGCGCGCTGACCGCGCATG 676
DB 612 lIeSerThrHlaLeuSerAlaSerAlaArgCysProThrAlaAlaArgSerGlyThrSer 631
QY 677 GCGCATCTACG-----CCACCATGCGCGACGAGAGACGACATCGTGGCG 727
DB |||||
QY 632 AlaValThrThrValThrCysThrThrProThrTrpSerSerSerThrSerAlaArgTrp 650
DB |||||
QY 728 CGTGGAGAGAACCTTCGCGCATCGCGCGCGCGCGCTG----- 763
DB |||||
QY 651 ArgThrArgSerProAlaSerArgTrpLysThrPmetProSerSerMetLeuLeu 670
QY 764 ---ACGTGCGCGTGTGATCTCGCACCAAGGTCATGCGGACGACCATTCGCGCGCT 820
DB |||||
QY 671 SerSerThrThrTrpGlnAlaArgThrArgAlaAlaSerTrpSerProLeuGlyLeuAla 690
DB |||||
QY 821 CGC-----GCGAGACGCTGCGCTGATGATGAGCGCGCA----- 853
DB |||||
QY 691 ArgSerLeuLeuProLeuAlaThrAlaSerProCysArg-----AlaThrProThrGly 708
DB |||||
QY 854 -----TGGCGCGCGACGACTCT----- 871
DB |||||
QY 709 SerGlyPro**ThrTrpArgSerCysSerSerTrpLysThrGluArgHlaArgAsnTrp 728
DB |||||
QY 872 -----CGTGGAGCGGT-----ATCCCTACG 892
DB |||||
QY 729 ArgGlnCysGlySerGlnGlySerAlaArgMetAlaArgThrArg**AlaAlaSer 748
DB |||||
QY 893 TGGCGCGCTCCACATGCTCAACGACGACCGCGCTGCTGCGCGGACGACGACATATCA 952
DB |||||
QY 749 TrpThrSerThrThrTrpLysAlaSerSerThrCysCysTrp----- 762
DB |||||
QY 953 CCTGTGCAAGCCTTCCCGCAACTGAGGCGCGGACCTGATGATGAGTGGCGCGGAC 1012
DB |||||
QY 763 -----TrrProThrLysTrpProCys 769
DB |||||
QY 1013 GCGGAAATCCAGTAGACAGTGTGCGCG---ACGTGACGCGCGCGCGCGCATCTACT 1069
DB |||||
QY 770 TrpSerSerProGlySerThrTrpSerThrGlySerCysAlaThrArgCysPro----- 787
DB |||||
QY 1070 TCATGATGAGGAAACCGACGCTGACGCGCATCTGCGCTGCGCGCGGACATGATGCGGT 1129
DB |||||
QY 788 -----ThrHlaProSerTrpThrSerCysTrpLeuSerAlaGlyAlaSerThrAla 804
DB |||||
QY 1130 CCGAGCGC---TGGCGACGACGAGCGCGCGCATCCGCGCTTGGGACACTTCGCG 1186
DB |||||
QY 805 AlaSerAlaGlyCysAlaGlyAlaSerProAlaHis-----ArgGlyArgProAlaArg 821
DB |||||
QY 1187 GGGGCTGGGACATATGCGCGCGCGCTGCGCTGCGCGTCCGCGTACAGCGCGTATAGA 1246
DB |||||
QY 822 ThrSer---ArgProAlaArgProArgProAlaCysSerArgPheCysArgGln----- 838
DB |||||
QY 1247 AGATGACCGCGC---TACCGCGCGCGCGCTTGGCGCGCGCGC----- 1288
DB |||||
QY 839 -----ProAlaThrTrp**ProArgArgAla**AlaThrProThrThrAlaProLeu 856
DB |||||

```

```

QY 1289 -----GCGGGCAGCTGACGGCGGGTACTTCCGCCGACTGCTGG 1327
Db 857 AlaProserArgIleGlyValAlaAlaAlaValAlaArgProHlaArgProAlaArgPro 876
QY 1328 TGTTCGACCGCCGACGGTGGCCGACCTTCCGAAACACCTTACCGAGCGCGCG 1387
Db 877 ArgIleValAlaProAlaHlaAlaCysProProHlaArgProGlnSerArgAlaPro 896
QY 1388 CCGGCATCCATTCGTGTACTGACAGCGCGCGCGGTGCGCAAGAGCAGCGCTTACCG 1447
Db 897 ArgAlaGlyAlaParGlnThrGlyValAlaAlaArgLeuGlyAlaGlyLeuArgSerPro 916
QY 1448 GCCAGCATGCGCGCGGTGCTCCGACGACGCGCGCTGAGCGCGCGCGCGCTTAC 1507
Db 917 ArgAlaAlaProArgArgArgGlyArgProCysProHlaSerProGlnCysArgAlaAla 936
QY 1508 AATCGCGCGTGAACGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1567
Db 937 GlnPro-----GlyArgArgGlyGlyArgGlyGlyProGlyThrAlaGlyGlyThr 953
QY 1568 TGGCCCTCCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1600
Db 954 SerArgProProSerGlyProCysArgProArgAlaValThrAlaProPheLeuGln 973
QY 1601 -----CCGATATGCTGGGCGCAAG 1618
Db 974 ProHlaSerProAlaAlaProSerSerArgSerSerArgSerProAlaGlyTTP----- 991
QY 1619 AATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1669
Db 992 -----ArgThrCysArgCysSerValArgSerSerTTPProGlyArgPro 1007
QY 1670 TGCAGACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1720
Db 1008 Cys**ThrArgProGlyProGlyAla--ArgAlaArgValThrLeuProCysProAla 1026
QY 1721 CCGAGCTGGGCGCGCA 1735
Db 1027 Pro--TTPProArg 1030

RESULT 8
AA47961
ID AA47961 standard; Protein; 1232 AA.
XX AA47961;
AC AA47961;
XX AA47961;
DI 05-MAR-2002 (first entry)
XX
DE Human NMDA receptor subunit SEQ ID NO 43.
XX
KW Human; N-methyl D-aspartate receptor; NMDAR; genetic screening;
KW glutamate receptor; drug; modulator; receptor.
XX
OS Homo sapiens.
XX
PN US6316611-B1.
XX
PD 13-NOV-2001.
XX
PF 29-SEP-1997; 97US-0940035.
XX
PR 20-APR-1994; 94US-0221193.
XX
PR 20-APR-1993; 93US-0052449.
XX
PA (MERI) MERCK & CO INC.
XX
PI Daggett LP, Lu C;
XX
WP1; 2002-096550/13.
XX
DR N-PSDB; AB199189.
XX
PT Novel nucleic acid encoding human N-methyl D-aspartate receptor 2A

```

```

PT subunit, useful for expressing the subunit and as probes in genetic
PT screening techniques, to identify patients having neuropathological
PT conditions
PS
XX Example 3; Column 257-264; 206pp; English.
CC The invention relates to nucleic acids (I, AA199167-AA199196, ABN05910)
CC encoding human N-methyl D-aspartate receptor (NMDAR) subunits. The
CC nucleic acids are useful for the production of NMDA receptor subunit
CC proteins (ABN56503-ABN56530, AA47961), as probes and to identify and
CC isolate nucleic acids encoding related receptor subunits. Genetic
CC screening can be carried out using the nucleotide sequences as probes to
CC determine if any abnormalities exist with respect to any of the
CC endogenous glutamate receptors. Also, patients having a family history of
CC disease states related to glutamate receptor dysfunction can be screened
CC to determine if they are also predisposed to such disease states.
CC Expression vectors comprising (1), receptor subunits encoded by (1) and
CC host cells comprising (1) are useful for producing selected NMDAR
CC subunits and in specific combinations, as well as antibodies to the
CC receptor subunits substantially free of contamination from many other
CC NMDAR subtypes. The availability of desired receptor subtypes allows the
CC observation of the effect of a drug substance on a particular receptor
CC subtype or combination of NMDAR subunits and to perform initial in vitro
CC screening of drugs or modulators in a test system that is specific for
CC humans, a human NMDAR subtype or the combination of NMDAR subunits.
XX
SQ Sequence 1232 AA;
XX
Alignment Scores:
Pred. No.: 5,32e-19 Length: 1232
Score: 378.50 Matches: 189
Percent Similarity: 33.15% Conservative: 58
Best Local Similarity: 25.37% Mismatches: 233
Query Match: 11.47% Indels: 265
DB: 23 Gaps: 44
US-10-009-782-1 (1-1758) x AA47961 (1-1232)
QY 32 CCATGTCC-----AATCGATP-----CCGACG 55
Db 374 ProCysProHlaAlaGlyArgAlaThrThrProSerAlaAlaGlyThrTTPProHla 393
QY 56 CCTGACCTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 112
Db 394 ProSerSerAlaValArgSerAlaSerThrSerSerArgSerTTPProGlnTTPSer 413
QY 113 GCGCGCGCG-----ACCTG-----GCGTGGCGGCG 139
Db 414 AsnSerProHlaThrCysThrTTP**ProHlaAlaSerMetAlaSerGlyCysAlaAla 433
QY 140 ACCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 199
Db 434 TyrGlyThrAla**LeuGlyArgCysThrThrSerGlyGlnThr----- 448
QY 200 CGGCGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 256
Db 449 -----TTPProSerAlaProSerProSerMetArgAlaAlaProArgSer**ThrSer 466
QY 257 TGTCTAGGCGTGCAGCATGACGCCAGATCTGCGAGCGCGTCAACAGGTGTACGG 316
Db 467 Leu-----TyrProLeuThrPargAlaAlaSerVal**TTP----- 478
QY 317 GCAATTTGGGCGCATGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
Db 479 -----TTPLeuAlaAlaMetAlaProSerProProArg 489
QY 374 ACGTCTGAGAGAGCGCGCTTACCGTTTTCGAGCGCTTGCAGCTACTGACCGGT 433
Db 490 ProSerTTPSerHlaAlaLeu-----Gln 498
QY 434 TGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493
|||||

```

Db 499 CysGly-----CysteusercysalaserleuTriPPro-----Ser 513
 QY 494 CCGCGGTCATGCGGACTGACAGCGCGCCACGAGAGAAATCGCGGCA----- 547
 Db 514 ProSerSerCysSerSerThrSerAlaSerAlaThrThrArgThrSerProGluAla 533
 QY 548 -----TCGCGACCTGG----- 559
 Db 534 ArgLeuSerLeuSerAlaSerProCysGlyArgTrpSerSerThrThrGln 553
 QY 560 -----CCGAGAGAGCATAG-----CCAGCG 580
 Db 554 CysProSerArgTrpArgGlyAlaProProAlaArgSerTrpPheTrpSerGlyProSer 573
 QY 581 GCGCCATGCGCATTTGACGCGCGCTTACCGCCCG-----CCGCGCGCGCACCA 634
 Db 574 LeuLeuSerSer-----SerSerProAspThrArgProThrTrpProProSer***Ser 591
 QY 635 CCGAGAGAG----- 643
 Db 592 LysSerAsnThrSerThrLeuSerArgAlaSerValThrArgSerPheSerGlyLeuLys 611
 QY 644 -----TCATGAGGTGT-----GCCGCGCGCTGAGCGCGCATG 676
 Db 612 IleSerThrHisLeuSerAlaSerAlaArgCysProThrAlaAlaArgSerGlyThrSer 631
 QY 677 GCGGCATCTAG-----CCAGCCACATGCGCGACGAGCGACACATCGGCGCG 727
 Db 632 AlaValThrThrValThrCysThrProThrTrpSerSerSerThrSerAlaArgTrp--- 650
 QY 728 CGGTGAGGAAACCTCCCGCATCGCGCGCGAGCTGG----- 763
 Db 651 ArgThrArgSerProAlaSerArgTrpGlySerTrpMetProSerSerMetLeuLeu 670
 QY 764 ---ACGTGCGGTGTGATGCTGCGACACAGATCATGGCGACCAATTCGCGCGCT 820
 Db 671 SerSerThrThrTrpGlnAlaArgThrArgAlaAlaSerTrpSerProLeuGlyLeuAla 690
 QY 821 GCG-----GCGAGACGCTGCGCGCTGATGAGGCCCGCA----- 853
 Db 691 ArgSerLeuLeuProLeuAlaThrAlaSerProCysArg-----ArgThrProThrGly 708
 QY 854 -----TGCGCGCGCGAGACGCT----- 871
 Db 709 SerGlyPro***ThrTrpArgSerCysSerSerTrpGlyThrGlnArgAsnTrp 728
 QY 872 -----CGTGTGAGCGGT-----ATCCTACG 892
 Db 729 ArgGlnCysGlySerGlnGlySerAlaArgMetArgTrpArg***AlaAlaSer 748
 QY 893 TGGCGGCTCCACCATCTCAGACGAGCGCGTGTGCTGCGCGGACGACCATCATCA 952
 Db 749 TrpThrSerThrThrTrpGlnAlaSerSerThrCysGlyTrp----- 762
 QY 953 CCGTGTGCAAGCCCTCCCGACACTGAGGGGCGGCGACCTGGATGAAGTGGCGCGGAGC 1012
 Db 763 -----TrpProTrpGlyTrpProCys 769
 QY 1013 GCGGCAATCAAGTACAGCTGCTGCGCG---AGCTGACGCGCGCGCGCATCTACT 1069
 Db 770 TrpSerSerProGlySerThrTrpSerThrGlySerCysAlaThrArgCysPro----- 787
 QY 1070 TCATGATGAGAGACCGACGTCAGCGCATCTGCGCTGCGCGCGACCATGATCGGT 1129
 Db 788 -----ThHisProSerTrpThrSerCysTrpLeuSerAlaGlyAlaSerThrAla 804
 QY 1130 CCGAGCGGC---TGCGGACAGACGAGCGCGCATCCGCGCTGTGGGACCTTCCCGC 1186
 Db 805 AlaSerAlaGlyCysArgAlaSerProAlaHis-----ArgGlyArgProAlaArg 821
 QY 1187 GGGTCTGTGGGACATAGCGCGGACCTGCTGTGCTGCGCGGAGACGCGGTATGA 1246
 Db 822 ThrSer---ArgProAlaArgProArgProAlaCysSerArgPheCysArgGln----- 838

QY 1247 AGATACCGGCGC-----TCACCGCGCGCGCTTGCGCTGCGCGCGCGC----- 1288
 Db 839 -----ProAlaThrTrp***ProArgArgAla***AlaThrProTrpThrAlaProLeu 856
 QY 1289 -----GCGGAGAGCTGACAGCGCGCGGTACTTGGCGACCTGGTG 1327
 Db 857 AlaProSerArgGlyGlyValAlaAlaAlaValAlaArgProHisArgProProAlaArgPro 876
 QY 1328 TGTTGACCGCGCGCGCGCGGTGGCGGATACCGCATCTTCGAAACCCCTACCGAGCGCGCG 1387
 Db 877 ArgGlyLeuAlaProAlaHisAlaCysProProProThrArgProGlnSerArgAlaPro 896
 QY 1388 CCGCATTCATCTCGTGTACGTACAGCGCGCGCGCGGTGGCGAGAGCAGCGGCTTCACCG 1447
 Db 897 ArgAlaGlyAspArgGlnThrGlyValAlaAlaArgLeuSerGlyAlaGlyLeuArgSerPro 916
 QY 1448 GCCAGCATCCGCGCGCGCGCTGCTGCGACGACGCGCGCGCTGACCGCGCGCGCGCTTAC 1507
 Db 917 ArgAlaAlaProArgArgArgGlyArgProCysProThrSerProGlnCysArgAlaAla 1536
 QY 1508 AATCCGCGCGTGAACGGGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1567
 Db 937 GlnPro-----GlyArgArgGlyGlyArgCysGlyProGlyThrAlaGlyGlyThr 953
 QY 1568 TGCGCCCTCCCTCCGCTGCGCAATACGCGCGCGC----- 1600
 Db 954 SerArgProProSerGlyProCysArgProAlaValThrThrAlaProPheLeuGlu 973
 QY 1601 -----CCGATATGCTGGCGAAG 1618
 Db 974 ProThrAspProAlaAlaProSerSerArgSerSerArgSerProAlaGlySerTrp----- 991
 QY 1619 AAGTATGAGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1669
 Db 992 -----ArgThrCysArgCysSerValArgSerSerTrpProGlyArgArgPro 1007
 QY 1670 TGCAAGACCTGTGCGACGAGCGCGCATCGCGGTCTGCA-----CCCTTCCAGAG 1720
 Db 1008 Cys***ThrArgProGlyProGlyAla---ArgAlaArgValThrLeuProCysProAla 1026
 QY 1721 CCGAGCTGGCGCGAGA 1735
 Db 1027 Pro---TrpProArg 1030
 RESULT 9
 AAB26240
 ID AAB26240 standard; Protein; 1081 AA.
 AC AAB26240;
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Human N-methyl-D-aspartate receptor subunit NMDAR1A #18.
 XX
 KW Human; N-methyl-D-aspartate receptor; NMDA; NMDAR1A; Ionotropic;
 KW glutamate receptor; drug screening; animal model; disease diagnosis;
 XX genetic screening.
 OS Homo sapiens.
 XX
 FH
 Key location/Qualifiers
 FT Misc-difference 116
 FT /label- unknown
 FT /note- "encoded by TGA"
 FT Misc-difference 140
 FT /label- unknown
 FT /note- "encoded by TGA"
 FT Misc-difference 162
 FT /label- unknown
 FT /note- "encoded by TGA"
 FT Misc-difference 183
 FT /label- unknown

CC	research to identify other, similar proteins. They can also be used as
CC	probes, for example in genetic screening, and in drug screening, as well
CC	as enabling the production of animal disease models.
xx	
sq	Sequence 1081 AA:
Alignment Scores:	
Pred. No.:	2,11e-18
Score:	370.00
Percent Similarity:	32.63%
Best Local Similarity:	24.97%
Query Match:	11,228
DB:	21
	Gaps: 43
US-10-009-782-1 (1-1758) x AAB26240 (1-1081)	
QY	32 CCATGTCCTC-----AATCGGANT-----CCCAGC 55
Dd	
Dd	210 ProcyserAlaIalaaIyAlrGalArThrThrProSerAlaIalaaIyThrTrpProPiothr 229
QY	56 CCTTGCACCTGTGCTGGCGGGCGGCCACTTCATGAGCGGACGACACACCCCG--GGC 112
Dd	
Dd	230 ProSerSerAlaValArgAspSerAlaSerThrSerSerAlaGlySerTrpProGIutPsr 249
QY	113 GGGGGCGCG-----ACTCG-----ECGTGCGGCGG 139
Dd	
Dd	250 AsnSerProThrThrcysThrTrp***ProThrAlaSerMetAlaSerGlycysAlaIa 269
QY	140 ACCGATCGCCGCAATCGCGCATCTGTGAGCGGGCGGCGCACACCCGGTGACGT 199
Dd	:::
Dd	270 TyrGIyAlaIa***LeucIyArgcysThrThrserylgnThr----- 284
QY	200 CGGGCTGTGTGTCGGCCCGGCTTCATGACGTGC---ACACCAGACGACAATACTAC 256
Dd	
Dd	285 -----TrpProSerAlaProSerProSerMetArgSnaIaProArgser***Thrs 302
QY	257 TGCTCAGGGGTGCGACATGAGCGCCAAAGATCTGCCAGGGCGTCAACAGGTGTACGG 316
Dd	303 Leu-----TyrProleuTrpTrgTrgAlaSerAla***Trp----- 314
QY	317 GCMAATTGGCGCATGACCTGCGCGCGCTGGCGGACGCAACCGCGCG--CCCGCTGG 373
Dd	315 -----TrpleuAlaIaMetAlaProSerProProArg 325
QY	374 ACCTGCTGAGAGAAGGGCGCTTACCGGTTCAGAGCGCTTCGCCGACTACCTGAGCGGT 433
Dd	
Dd	326 ProSerTrpSerHisIleAlaLeu-----GlN 334
QY	434 TEGGGGCAACGCGCGCGCGCTCAACGCGCGCTGTATGGTGGGCCATTCACGCTGGCGG 493
Dd	
Dd	335 Cysgly*****CysLeuSerCysAlaSerLeuTrpTrpPro-----Ser 349
QY	494 CCGGCGTCATCCCGGACTTCGACGCGCGCGCCACCGACGAGAMAATCGCGCCA----- 547
Dd	
Dd	350 ProSerSerCysSerThrSerAlaLeuSerAlaIaThrThrArgThrSerProGIuaIa 369
QY	548 -----TGGGGAGCTCG 559
Dd	370 ArgSerProGIyAlaGlnLeuSerLeuSerAlaIaSerProCysglyCysCysglyArgTrp 389
QY	560 -----CCGAGAGAACCCATGG----- 574
Dd	390 SerSerThrThrGlnCysProSerArgThrArgGLyAlaProProAlaArgSerTrpHe 409
QY	575 -----CCAGGGGGCCATCGGATTTGCACCGGGCGCTTCACCCGCG-- 619
Dd	410 TrpSerGlyProSerLeuLeuSerSer-----SerSerProAspTrHrArgProThrTrp 427
QY	620 CCGCGCGCGCGCACCAACGAGAGA----- 643
Dd	428 ProProSer***SerIysSerAsnThrSerThrLeuCysArgAlaSerValThrArgSer 447
QY	644 -----

Db 774 ArgProCysPThrSerProGluCysArgAlaIaIaGlnPro-----GlyArgArg 799
 QY 1532 GCGCGCCCTCCCAACCTCGAGCGCAAAACGGTATATGCCCCCTCCGCTGCGCATATA 159
 Db 791 GlyGlyArgCysGlyProGlyThrAlaGlyLeuThrSerArgProProSerGlyProCys 810
 QY 1592 GCGCCCGCAC-----160
 Db 811 ArgProArgAlaValThrThrAlaProPheLeuGluProThrAspProAlaIaIaProSer 830
 QY 1601 -----CGATATCGTGGGCGCAAGAGTANGGGCGCGCGCTCGCGC 164
 Db 831 SerArgSerSerArgSerProArgSerTyr-----ArgThrCysArg 844
 QY 1643 CCGAGCCGACAG-----CCGCGAATAATGACCCCGAAGACCTGTCCAGCCACGCG 169
 Db 845 CysSerValAlaArgSerTyrProGlyGlyArgProCys***ThrArgProGlyProGly 864
 QY 1694 GCATCGGGGCTGCA-----CCCTGTCCAGAGCGAGCGAGCCGACAGA 1735
 Db 865 Ala---ArgAlaArgValThrLeuProCysProAlaPro---TyrProArg 879

05-MAR-2002 (first entry)

Human NMDA receptor subunit SEQ ID NO 44.

KW glutamate receptor; drug; modulator; receptor.
XX

Homo sapiens

PN US6316611-B1
XY

PD 13-NOV-2001.
yy

29-SEP-1997; 97US-0940035

20-APR-1994; 94US-0231193

XXX
XXX
/MEND , 1988-1989

XXXXXXXXXXXXXXXXXXXX

WPB. 2000 000000 0000

N-PSDB; ABA05910.

Novel nucleic acid

screening technology

Example 3: Conjecture

The invention of

encoding human N-1

proteins (ABB5650).
isolate n1c1a1c 2a

determine if any screening can be

endogenous glutamate disease states rat

Expression vectors

most cells comprise subunits and in sm

CC receptor subunits substantially free of contamination from many other
CC receptor proteins whose presence can interfere with analysis of a single
CC NMDAR subtype. The availability of a desired receptor subtype allows the
CC observation of the effect of a drug substance on a particular receptor
CC subtype or combination of NMDAR subunits and to perform initial in vitro
CC screening of drugs or modulators in a test system that is specific for
CC humans, a human NMDAR subtype or the combination of NMDAR subunits.
CX

Sequence : 1081 AA;

Alignment Scores:

File: NO.:	2.11e-18	Length:	1083
Score:	370.00	Matches:	189
Percent Similarity:	33.63%	Conservative:	58
Best Local Similarity:	24.97%	Mismatches:	234
Query Match:	11.22%	Indels:	276
DB:	23	Gaps:	43

00 20 003-102-1 (1-1/58) X ABB56523 (1-1081)

[illegible]

Db 428 ProPaser***SerLysSerAsnThrSerThrLeuCyAArgAlaSerValThrArgSer 447
 |||||
 Oy 644 -----TCATCGAGGTGT-----GCCGCC 661
 Db 448 PheSerGlyLeuLysIleSerThrHisLeuSerAlaSerAlaArgCyProThrAlaAla 467
 Oy 662 CGCTGAGCGGCATAGCGCGGCTATACGACCCACACATGGCGGACGAGACGACATCG 721
 |||||
 Db 468 ArgSerGlyThrSerAlaValThrThrValThrCysThrProThrTyrPseSerSerThr 487
 Oy 722 TGGCGCGGCTGGAG-----AACTCCGCATCG 751
 |||||
 Db 488 SerAlaThrThrPheThrArgSerProAlaSerArgTrpAlaLeuArgLeuSerLeuSer 507
 Oy 752 GCCCGGAGCTGG-----ACGTCCGCTGTGATCTCG 784
 |||||
 Db 508 ProGlySerThrPheMetProSerSerMetMetLeuSerSerThrThrTyrGlnAlaArg 527
 Oy 785 ACCGCAAGGTCATAGGCCACGCCCAATTCGGCGGCTGC----- 823
 |||||
 Db 528 ThrArgAlaAlaSerThrPseProLeuGlyLeuAlaArgSerLeuLeuProLeuAlaThr 547
 Oy 824 CGGAGACGCTGGCGCGGATCGAGCGGCCCA-----TGGCGCGCC 862
 |||||
 Db 548 AlaSerProCyAArg-----ArgThrProThrGlySerGlyPro**ThrTrpArgSer 565
 Oy 863 AGGACGCTC----- 871
 |||||
 Db 566 CySerSerThrGlyThrGluArgHisArgAsnThrPArgGlnCysGlySerGlnLys 585
 Oy 872 -----CGCTGAGCGGT-----ATCCCTACGTGGCGGCTCCACCATGCTCAAGC 916
 |||||
 Db 586 AlaArgMetArgThrArg***AlaAlaSerThrPheSerThrThrThrThrPheGluAla 605
 Oy 917 AGGACGCGCTGCTGCTGGCGCGGACGACATCATCACTGGTGTCAAGCCCTTCCCGAAC 976
 |||||
 Db 606 SerSerThrCysCyAArg----- 611
 Oy 977 TGAAGCGGCGGACCTGGATGACTCGGCGCGACGCGGCAATTCAGATGACAGCTGG 1036
 |||||
 Db 612 -----TrrProThrGlyTyrProCysTyrPseSerProGlySerThrThr 626
 Oy 1037 TGGCCG---ACGTGACGCGGCGCGGCCCATCTACTTCATGATGACAGAACCCAGCTGC 1093
 |||||
 Db 627 SerThrGlySerCysAlaThrThrCysPro-----ThrHisProSerThr 641
 Oy 1094 AGCGCATCTGGCTTGGCGCGACCATGATCGCTCCGCGACGCG---TGGCGACAGCG 1150
 |||||
 Db 642 ThrSerCysTrpLeuSerAlaGlyAlaSerThrAlaAlaSerAlaGlyCysAlaAlaSer 661
 Oy 1151 AGCGCGCGCATCGCGGCTGGGCGACCTTCGCGGCGGTCTGGGCGACATGACGCGCG 1210
 |||||
 Db 662 ProAlaHis-----ArgGlyArgProAlaAlaGlnThrSer---ArgProAlaArgPro 677
 Oy 1211 ACCTGGGCTGTCCGCTGGAGACGCGCGGTATGAAGTGAACGCGGC-----TGACCG 1264
 |||||
 Db 678 ArgProAlaCysSerAlaArgPheCysArgGln-----ProAlaThrTrp**Pro 693
 Oy 1265 CCGGCGGCTTCGGCTGGCGCGGC-----GCG 1291
 |||||
 Db 694 ArgTrpAla**AlaThrProThrThrAlaProLeuAlaProSerArgIleGlyValAla 713
 Oy 1292 GCGACGCGACGCGGCGGTACTTCGCGACGCTGGGTGTTCGACCGCGGCGAGCGTGGCG 1351
 |||||
 Db 714 AlaAlaValArgProHisArgProThrAlaArgProArgGlyLeuAlaProAlaHisAla 733
 Oy 1352 ATACCGGCACTTCGAACAACCTTACCGACGCGCGCGCGGACATCATTCGCTGACGTGA 1411
 |||||
 Db 734 CyArgProProThrThrArgProGlnSerArgGlnAlaThrArgAlaGlyAspArgGlnThrGly 753
 Oy 1412 ACGGCGCGCGGTCTGGCAAGGCGGCTTCAACCGGCGACATGCCGCGCGGCTGTCTCG 1471
 |||||

Db 754 ValAlaArgLeuGlyLeuArgSerProAlaAlaProAlaArgArgGly 773
 QY 1472 CACGACGCGCGCTGAGCCCGGCGAAGCCCTTACATCCGCGGAGCGGCGCGT 1531
 Db 774 ArgProCysProThrSerProGluCysAlaGlyAlaGlyPro-----GlyArgArg 790
 QY 1532 GCGGCCCCCGCCACCCCTGAGCAACCGTACATGGCCCTCCCTCCGCTCGCAATA 1591
 Db 791 G1G1ArgGlyGlyProGlyThrAlaGlyGlyThrSerArgProProSerGlyProCys 810
 QY 1592 CCGCCCCCAG-----CCGATATCGTGGGCAAGAGTATGGCGCGCGCTGCGCG 1600
 Db 811 ArgProArgAlaValThrThrAlaProPheLeuGluProThrAspProAlaAlaProSer 830
 QY 1601 -----CCGATATCGTGGGCAAGAGTATGGCGCGCGCTGCGCG 1642
 Db 831 SerArgSerSerArgSerProArgSerTrp-----ArgThrCysArg 844
 QY 1643 CCGACGCGAAG-----CCGGAATAATGACCTGCAAGACCTGCGAGCGCGCG 1693
 Db 845 CysSerValArgSerSerTrpProGlyGlyArgProCys**ThrArgProGlyProGly 1735
 QY 1694 GCATCGCGGTCTCGA-----CCCTGTCAAGCGCGAGCTGGCGCGAGA 1735
 Db 865 Ala--ArgAlaArgValThrLeuProCysProAlaPro--TrpProArg 879

RESULT 11
 ABG03533
 ID ABG03533 standard; Protein; 900 AA.
 AC ABG03533;
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #3524.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.

XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 PA (HYSER-) HYSER INC.
 XX Drmanac RT, Liu C, Tang YF;
 DR N-PSDB; AAS67720.
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS Claim 20; SEQ ID No 33892; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC at fip.wipo.int/pub/published_pot_sequences.

XX Sequence 900 AA;

Alignment Scores:

Pred. No.: 2,5e-16
 Score: 341.00 Length: 900
 Percent Similarity: 32.59% Matches: 203
 Best Local Similarity: 28.16% Conservative: 32
 Query Match: 10.51% Mismatches: 214
 DB: 22 Gaps: 272
 Indels: 41

US-10-009-782-1 (1-1758) x ABG03533 (1-900)

QY 1750 GGTACCTCAGGCGGCGATCTGCGCCAGCTGCGCTTGAGAGGCTGACCGCGATGCCGC 1691
 Db 9 ArgSerAlaArgSerValGlyArgArgPro-----GlyThrAlaProAlaAlaArg 26
 QY 1690 TGGCTGCGACAGCTTTCAGGCTCATTTCCGGCTTCCGCTGCGCGCGCGCGCG 1631
 Db 27 ProPro-----ArgProAlaAla 32
 QY 1630 CCGCCATCACTCTCTCCACGATATGCGGTGGCGCGATATGCGAGCGAGGAGG 1571
 Db 33 GlnArgGlnAlaLeuGlySerArgGlyArgValGly-----ThrGly 46
 QY 1570 CCATGACGCGCTTTCGCTCCAGG-----TTGGAGGCGGCGCGCGCGCG 1526
 Db 47 ProGlyArgGlyLeuArgProGlyGlyTyrGlyCysPheProGlyProArgGlyThrGlu 66
 QY 1525 CCGCGTTCACGCGCGATGTAAG-----GCTGGCGCGCGCGCTCAGCGCG-----1481
 Db 67 AspAlaAspGlnArgAlaAlaArgGlyProValGlyAlaGlyThrGlnHisGlyArg 86
 QY 1480 -----CGTGC 1475
 Db 87 AlaValProArgGlyProGlnAsnGluProAspGluThrLeuProGlyGlyProSer 106
 QY 1474 GTGCGAGCA-----CGGCGCGCGATCTGCGCGTGAAGCGCT---GCTTGC 1427
 Db 107 ProArgGlyGlyGlyLeuArgGlyArgSerGlyAlaArgGlyLeuPro**SerLeuThr 136
 QY 1426 AGACCGCGCGCGCTTGCAGTACACGGAATGATGCGCGCGCGCTCGTGGGTGT 1367
 Db 127 GlyProAla-----GTCGACACCA-----CCA 1322
 QY 1366 CGAAGTGGCGGTATGCGCCACCGTGC-----GTCGACACCA-----CCA 1322
 Db 130 -----ProGlyProGlnArgGlyGlyGly**SerProSerPro 142
 QY 1321 GGTGCGAGTACCGCGCGCTGACAGTCCGCGCGCGCGCGCGAGCGCGCGG 1262
 Db 143 GlyArgAlaSerSerArg**AlaGlyProTrpGlyArgProGlyAlaSerArgAlaSer 161
 QY 1261 TCAGCGCGGTATCTTCATACCGCGCGCTCCAGG-----1226
 Db 162 LeuGlnArgAlaSerSerMetProAlaSerGlnValAspTrpGlyGlyProGlyGlySer 181
 QY 1225 -----GGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1199
 Db 182 ProArgCysAsnArgCysArgGlyArgGlySerProGlyThrGlyProGlyTrpProArg 201

```

QY 1198 GCCCAGCACCAGCGGAGAGTGCCTC-----ACAGGCGGATGCGGGCGCT 1151
Db 202 LeuArgSerProGlyAlaSerLeuArgProGlyValGlyGlyLeuGlyLeuAlaLeuProAla 221
QY 1150 CGTGTGTGGGAGAGCCGTCGAGACCGATCATGTGCGGCGCAACCCAGATGCGCTGCA 1091
Db 222 ArgThrAlaAla-----ProGluProAlaAlaProAlaArgGluPro 525
QY 1090 CGTGTGTGTGTCCATCATGATGATGCGCGCGCGCGCTGCA-----GCT 1043
Db 228 ProArgProArg-----GluArgTrpArgSerProGlyAlaProCysLeuGlyAla 244
QY 1042 CGGCGACACCGCTGTACTTGTGATTTGCGCGCTGCGCGCGCATCTCATCAGATGCGCGC 983
Db 245 Gln**Pro-----SerLeuProAlaAlaAlaGluProArg---GlyProGlyThrSer 261
QY 982 CGCTCAGTGTGGGAGAGGCTGCACACGATGATGATG----- 944
Db 262 -----GlyArgAla-----TrpProSerSerAlaAlaAlaArg 272
QY 943 -----TGCCTCCGCGCAGCAGCAGCGCTCTGTGAGCATGTTGAGCGCG----- 896
Db 273 GluArgPys---ArgAlaProGlyArgGlyProAlaAlaProThrGlyAlaArgArgPro 291
QY 895 CCAAGT-----AGGATACGCGCTCCAGCAGCAGCTGCGC----- 860
Db 292 ProArgProGlyAlaGlyLeuAlaGlyArgProThrArgAlaAlaLeuProGlyArg 311
QY 859 -----GGCCATGCGCG-----CCTCGATCAGCGCGCAGCGCTCGCGCAGC 818
Db 312 IleGlyArgArgSerProGlyArgAlaAlaProHisSerProAlaGlyLeuArgSerAla 331
QY 817 GCGCGAATTTGGCTGCGCATGACCTTGTGTGCGAGATCACCACG----- 770
Db 332 AlaArgArgGysProProProProProAlaProArgGlyGlnProGlyLeuProArg 351
QY 769 -----GCACGTCCAGCTCGC 755
Db 352 GlyGlyGlyAlaAlaAlaProLeuArgHisGlyGlyGlyAlaGlnLysThrArgProSerGln 371
QY 754 GCGCGATGCGGAGAGTTCCTCCAGCGCGCGCGCGAGATGCTCTCGCTCGCGCATGT 695
Db 372 GlyArgProGlyProProProProProAlaProAlaGlyGlnArgGlnProGly 391
QY 694 GGGTGGCGTATGCGCGCATGCGCGCTCAGCGCGCGCAGCAGCTGATCTCTCGG 635
Db 392 LysGluSerGlyGlnLysGlyValGlnProAlaArgGlyGlnProGlnGlyProGlnArg 411
QY 634 TGTGTGCGGCGCGCGCGCGCGGTAGAGCGCGCGCTCGAATGCGGATGCGCGCTG 575
Db 412 AlaMetArgGlyAlaAlaProAlaGlyArgArgPro----- 423
QY 574 CCATGCGCTTCCTCGCGCAGGTCGCGCATGCGCGCATTTCTCGTGGTGGCGCGCGCT 515
Db 424 -----ProAlaProAlaProGlyAlaAlaAlaGlnProGlyArgTrpGlyThrAla 440
QY 514 GCAAGTCCGCGCATGACCGCGCGCGCGCGCGTGAATGCGCCACCATACAGCGCGCTGA 455
Db 441 AlaGlyAlaAlaGlyPro-----AlaAlaGlyArgProAlaProGlyArgGlnProSer 458
QY 454 CGGCGCGCGCGCGCGCGCGCGCGCATGCTGCGCGCGAGCGCTCGAAGCTGAAG 395
Db 459 ArgAlaAlaGlnGlyProAlaAlaAlaAlaProGlyGluArg-----GlyGln 473
QY 394 AGC-----CGCCTTCGCGCAGCAGCGCTCAGGCGG 365
Db 474 SerLeuProValProIleGlyValAlaGlnGluArgGlyGlyProArgGlyPro 493
QY 364 CGGCGCGGTGTGGCGCGCGCGCGCGCGCGCATGCGCGCAATTTGCCCGGACACCG 305
Db 494 ThrAlaAla---AlaGlyProAlaAlaProGly----- 503
QY 304 TGTGTAGCCCTCGCAGATCTTGGCGGTCAATGTGCGCAGCGCTGAGCAGTGTGTGT 245

```

```

Db 504 ----- 512
QY 244 ---CGTGGGTGTGCGATGATGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 188
Db 513 AlaArgGly-----ProGluProAlaAlaProAlaArgGluPro 525
QY 187 GGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155
Db 526 GlyArgArgAlaAlaGluProAlaArgThrArgGlyProProAlaProAlaProSerGluPro 545
QY 154 TGGCGCGCATGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104
Db 546 TrpProLeuProSerArgProProGlySerGlyLysArgGlyAlaSerAlaProProGly 565
QY 103 TGTGTGCTCGCTCGA-----TGAGG 83
Db 566 GlyCysProSerArgSerGlyAlaSerProThrLeuGlnAlaProGlyProHisSerGly 585
QY 82 TGCAGC-----CCGCGAGCAGCAGT-----CGAAGGCTGGGAATCGGATGG 38
Db 586 CysProGluGlyPheProAlaArgLeuAlaGlyGlyGluArgSerSerGlyProGluMetCyl 605
QY 37 ACA 35
Db 606 Thr 606

RESULT 12
AAB59827 standard; Protein; 1592 AA.
AC AAB59827;
AC AAB59827;
DT 04-APR-2001 (first entry)
DE Protein #4 encoded by TuteD/E gene.
DE
KW Toluene degradation; enzyme; waste degradation; Tute; TuteD.
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azarcus toluilyticus.
XX
PN MO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000MO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNITV OHIO.
XX
PI Coschignano PW;
XX
DR WPI: 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition:toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tute (see AAF23628 and AAB59831), tute (AAF23630 and AAB59832),
CC tute (AAF23631 and AAB59833) and tute (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TuteD/E.

```

152 seragspargserargalaserhyserargglyser-----ProLeucysglyala 809

1128 CysAlaGlnHisAsnAlaGlnGlnTyrGlnValSerAlaGlnGln...Probnh...
||||| . ||||| |||

DB 614 rProtrHyTPProAlaSerSerCylSerProCylAlaSerAlaSerThProAl 634
Y 1600 CCCGATATGTGGGCAAGAAAGTGAATG---GCGCGCGCTTCGCGCCAGGCAAGCC 1656
Db 634 aserAlaSerTPAlaHisSerArgPheArgSerSerThrCysSerSerTysAlaArgAs 654
Y 1657 CGGAAATGACCTTCGACAGACCTGTGCGAGCCAGCCGATCGCGGTCTCGACCCCTGTC 1716
Db 654 pSerTyrArgAspHisGluThrHisArgCysProHisPheAlaArgGlnAlaAlaArgVa 674
Y 1717 AAGCCGACCTGGGCGACATGCTCCCTGACGACTACGA 1751
Db 674 lArg-SerAlaGlyHisAspArgAlaAlaArgLysArg 685

RESUTP 14
AAB59826
ID AAB59826 standard; Protein: 1615 AA.
XX AAB59826;
AC
XX
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #3 encoded by TufD/E gene.
XX
KM Toluene degradation; enzyme; waste degradation; TufD, TufE.
XX
XX Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azotarcus toluilyticus.
XX
XX WC200072650-A2.
XX
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000MO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
XX (UYOH-) UNITV OHIO.
XX
XX
PI Coschigano PW;
XX
DR WPI: 2001-041080/05.
XX N-PSDB; AAF23627.
XX

Composition comprising toluene degrading enzyme useful for biological treatment of organic compounds, especially for degrading toluene or its analogs -

Disclosure: Fig 12; 122pp; English.

The present invention relates to toluene degrading enzyme genes and proteins tufD (see AAF23629 and AAB59831), tufE (AAF23630 and AAB59832) CC CC tufF (AAF23631 and AAB59833) and tufG (AAF23632 and AAB59834). The CC CC toluene degrading enzymes are homologues of pyruvate formate lyase. The CC CC toluene degrading enzymes are useful for biological treatment of organic CC CC compounds and in particular for the degradation of toluene and its CC CC analogs contained in liquid or solid waste source. The present sequence is a protein sequence encoded by toluene degrading enzyme gene, tufD/E.

Seq Sequence 1615 AA;

Alignment Scores:
Pred. No.: 1.42e-15
Score: 331.50 Length: 1615
Percent Similarity: 36.65% Matches: 178
Best Local Similarity: 28.12% Conservative: 54
Query Match: 10.05% Mismatches: 203
DB: Gaps: 22 Indels: 128
36

PD 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 DR N-PSDB; AAS67717.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS Claim 20: SEQ ID No 33889, 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPRO
 CC at ftp.wipro.int/pub/published_pct_sequences.
 CC
 SQ Sequence 599 AA;

Alignment Scores:

Pred. No.: 5,74e-15 Length: 599
 Score: 321.50 Matches: 196
 Percent Similarity: 34.57% Conservative: 37
 Best Local Similarity: 29.08% Mismatches: 233
 Query Match: 9.75% Indels: 210
 DB: 22 Gaps: 37

US-10-009-782-1 (1-1758) x ABG03530 (1-599)

QY 53 AGCCCTGACGCTGCTCTCCGCGGCGACCTCTACGACGACCAACACCCGG--- 109
 DB 21 SerProAlaProGlyCysArgArgAlaAlaProAlaGlyTrpSerSerProGlyProAlaGlyTrp 40
 QY 110 -----GGCGGCGCGGCGACCTGCGGGGTGCGGCGGCGACCGCATCGCGGCGGATC 163
 DB 41 AlaAlaGly**ArgArgSerTrp-----CysAlaSerAlaSerLeuAla**SerPro 58
 QY 164 TGTGCGACGCGC-----CCGCGGACCA 184
 DB 59 CysArgProProAlaArgSerArgTrpTrpArgAspAlaGlySerGlyTrpThrProAlaCys 78
 QY 185 CCCGGGTGACGCGTGGCGCGCG-----TGTGCGGCG 217
 DB 79 ProAlaSerAla-----AlaTrpGlyAlaGluGlnGluProAlaArgSerTrpGlyPro 96
 QY 218 CCGGCTTCATGCACTCGACACACCC----- 241

DB 97 ArgAlaSerGlnSerHisCysProGlyGlyLeuAlaArgAlaProProGlySerValArg 116
 QY 242 ---ACGACGACAACTACTCTCTCAGGCGTGGCGACATGAGCCCAAGATCTCCAGGCG 298
 DB 117 CysSerThrGln**AspCysSerSerVal----- 126
 QY 299 TCACCAAGGTGTACAGGCGGATGCGGCATTCAGCTGGCGCGCTGGCGACGCCAAC 358
 DB 127 ArgProAlaTrpSerArgSer**GlyAlaCys**GlnVal**ProArgCysProCys 146
 QY 359 CGCCCGCCCGCTGACCTGCGGTGCGAGGAGGCGGCTTACCGTTTCAGCCCTTCGCG 418
 DB 147 ArgThrProAla---ThrGlyTrp-----AlaPro 155
 QY 419 ACTACTGACAGCGGTGCGGCGCGCGCGCGCG---TCAGCGCGCGCTGATAGTGG 475
 DB 156 ProProGlnGlyArgCysGlyProGlyTrpThrAlaProGlySerThrGlyProAlaGlyArg 175
 QY 476 GCCATTCAACGTGGCGGCGCGGCGGTCAAGCC----- 506
 DB 176 Ala---SerLeuCysCysProArg-ArgAlaHisLeuProGly**TrpProGlnLeuLe 194
 QY 507 -----GACTTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 535
 DB 194 uIleCysAlaHisProGlyAlaAlaSerLeuGlyLeuAlaCysGlnProHisAlaArgGlyLy 214
 QY 536 AAATGCGCGCCATGCGGCGGACCTGGCCGAGAACCATGCGCGCGCGCA-TGCGCAT 594
 DB 214 sGlyThrProIleGlnGluProAlaCysGlyTrp**GlyLeuArgArgGlySerGlyC 234
 QY 595 TCGACCGCGCGCTTACCGCG-----CCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645
 DB 234 sProGlyArgProHisThrAlaArgArgCys**ProProAlaProCysGlyArgArgse 254
 QY 646 ATCAGAGTGTCCGCGCGCGCGCATGCGCGCATGACCGCGCGCGCGCGCGCGCGCG 705
 DB 254 Ala-GlySerAlaHisProAlaArgProTrpHis-----GlyProGlyGlyGln 272
 QY 706 GAAGG-----CGAGCAATCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 756
 DB 272 InaArgAPProGlyProAlaArgGlyArgGlyGlnGlyArgSerProAlaSerPro 292
 QY 757 GAGCTGACGTGCGCGGCGGTGATCTGCACACCAAGTCTAGGCGCAATTTCCG 816
 DB 292 erGlyArgArgLeu-----P 297
 QY 817 CGCTCGCGGAGACGCTGCGCGTATCGAGGCGCGCATGCGCGCGCGCGCGCGCGCG 876
 DB 297 roAlaSerArg---AlaGlyArgSerArgAlaAlaArgGlyThrProGlyArgProGlu 316
 QY 877 GACGCGTATCCCTACGT-----GGCGGCGTC-----C 903
 DB 316 roArgSerProGlnArgAlaGlyThrGlyThrValGlnProAlaArgCysProTrpProPon 336
 QY 904 ACCATGCTCAACAGACGCGCGTGTGCGGCGG-GAGCGACATCATCATCTGTTGCA 962
 DB 336 IsArgAlaAlaAlaGlyProProAlaArgGlySerGlyAlaProAlaProLeuGlyArgT 356
 QY 963 GCCCTTCCCGAATCTGACGCGGCGCGAC-----TGGATGAAGCGCGCGCGCGCG 1016
 DB 356 hrArgSerPheGlyThrAlaGlyLysAlaHisProTrpProArgArgArgProGly---- 374
 QY 1017 CAATCAAGTACGACGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1076
 DB 375 -----HisTrp**SerAlaAlaAlaAlaProAlaThrGlyVal----- 387
 QY 1077 GAGACACCGCGACGTGCGCGCA---TCTGCGCGTGGCGCGCGCGCGCGCGCGCGCG 1133
 DB 388 -----ProAlaCysArgAlaGlySerTrpValSerAlaAlaProProAlaGlu-Gly 404
 QY 1134 CGGCGTGC-----GCACGACGAGCGCG-----GCATCGCGCGCGCGCGCG 1175
 DB 405 ArgProAlaArgAlaArgAlaArgHisProGlyArgCysProGlyLysAlaSerGlyProAlaGly 424

```
QY 1176 CACCTTCCGCG-----GGTGGGGGACATACCGCGACCTGGGCTGT 1223
Db 425 ArgArgSerAlaAlaHisGlyHisGlyAlaArgAlaGlySerProGlnProGlyAlaPro 444
QY 1224 CCGCGTGAAGACGCGGTATGAAAGATGACCGGCT-----GACGCGCGCGGCTT 1274
Db 445 ProGlyHisLeuProGlyLeuProAlaArgGlnProLeuGlyLeuProAlaArgGlnArg 464
QY 1275 CCGGCTTGGC-----CGGCGCGCGGACCTGACGCGCGGCTACTGCGCGACCT 1322
Db 465 CysPheGlyGlyLeuAlaGlnArgGlyArgAlaAlaArgHisCysLeuLeuSerArgPro 484
QY 1323 GGTGGTGT-----CGACCGCGCGACGCGTGGCGGATACGCG 1358
Db 485 SerAlaLeuAlaLeuArgHisSerSerTyrArgGlnProGlyMetGlyGly----- 501
QY 1359 CACCTTGAACACCTTACCGGACGCGCGCGGACATCCATCCGTGTACGTACGACGCGCG 1418
Db 502 -----TTPArgSerProGlnAlaLeuGlyGlyTyrGlyGlySerGln 516
QY 1419 GCCGCTTGGCAAGACGCGCTTACCGGCGGACATGCGCGCGCGGTGTCGACGCGAC 1478
Db 517 AlaGlySerAlaAlaGlyLeuSer-----GlyAlaAlaSerGln 528
QY 1479 GCCGCTTGAAGCGCGCGCGACCTTACAATCCGCGGTGAACGCGGCGCGGTGCGG--- 1535
Db 529 GlyArgArgAla-----ArgHisLeuArgGlyLysAlaProAla 541
QY 1536 -----CCCCCTCCACCTTGGACGCAACCGCTACATGCGCCCTCCCTCCGC 1583
Db 542 TrpAsnProAlaProProSerPro---ProProProAlaLeuGlyLeuProLeuArg 560
QY 1584 TCGCAATACGCGCGCGGATATCGTGGCAAGAGTGAATGGCGCGCGCTGCGCGC 1643
Db 561 ThrGlnArgGlnAlaThrArg-----LysProArgArgGlnGlnAlaArg 575
QY 1644 CGAGCGCAAGCGCGGAAATGACCTTGAACGCTTGGCAGGCGGCGGATCGCGGT 1703
Db 576 Arg-----ProArgProArgProLeuArgPro----- 584
QY 1704 CTCGACCTGTCCAAAGCGCGGATGCGCGGCGGCGGATCGC 1739
Db 585 ---GlyGlyAlaAsnGlySerProGlyProProArg 595
```

Search completed: May 11, 2003, 12:09:19
Job time : 97 secs

GenCore version 5.1.4.p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 01:22:49 ; Search time 28310 Seconds

(Without alignments)
1005.711 Million cell updates/sec

Title: US-10-009-782-1

Sequence: 1 gaattcacttgatcgagcga.....ccctgagctacgagaagctt 1758

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

EST.*

1:	em_estba.*
2:	em_esthum.*
3:	em_estlin.*
4:	em_estmu.*
5:	em_estov.*
6:	em_estpl.*
7:	em_estro.*
8:	em_hic.*
9:	gb_est1.*
10:	gb_est2.*
11:	gb_hic.*
12:	gb_est3.*
13:	gb_est4.*
14:	gb_est5.*
15:	em_estfun.*
16:	em_estom.*
17:	gb_gss.*
18:	em_gss_hum.*
19:	em_gss_inv.*
20:	em_gss_pln.*
21:	em_gss_vit.*
22:	em_gss_fun.*
23:	em_gss_mam.*
24:	em_gss_mus.*
25:	em_gss_others.*
26:	em_gss_pro.*
27:	em_gss_rtd.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	83	4.7	1516	12	BG809984 mgct002xd
2	82.8	4.7	1798	17	AG171124 Pan trogl
3	82	4.7	1798	17	AG171124 Pan trogl
4	78.4	4.5	1542	17	AG032943 Pan trogl
5	76	4.3	1189	17	AG030608 Pan trogl
6	74	4.2	1116	12	BG810038 mgct002xi

7	72.4	4.1	925	17	CNS0091P
8	72.4	4.1	1299	17	AG039481
9	71.8	4.1	932	17	CNS00720
10	71.8	4.1	1341	17	AG030611
11	71.6	4.1	925	17	CNS0091P
12	71.2	4.1	1152	17	AG076818
13	71.2	4.1	1538	17	AG030607
14	71.1	4.0	1129	17	AG127788
15	71.1	4.0	1281	12	BG852363
16	71.1	4.0	1288	14	BG678719
17	70.2	4.0	935	17	CNS006XK
18	70.2	4.0	1131	17	AG042920
19	69.6	4.0	1625	17	AG043477
20	69.2	3.9	1360	13	BM463178
21	69	3.9	1160	14	BM926731
22	68.6	3.9	935	17	CNS006XK
23	68.6	3.9	1364	14	BM810045
24	68.4	3.9	932	17	CNS00720
25	68.4	3.9	982	14	BG687717
26	68.4	3.9	1626	10	AW731212
27	68.2	3.9	1387	10	AW731212
28	67.8	3.9	936	12	BG852371
29	67.8	3.9	1232	17	AG072425
30	67.8	3.9	1328	17	AG043615
31	67.6	3.8	1088	17	AG075424
32	67.6	3.8	1299	17	AG039481
33	67.4	3.8	803	17	AG162326
34	67.4	3.8	1189	17	AG030608
35	67.4	3.8	1278	17	AG030608
36	67.4	3.8	1651	12	BG809816
37	67.2	3.8	644	12	BG607300
38	67.2	3.8	1248	17	AG049208
39	66.6	3.8	1040	12	BG786331
40	66.4	3.8	1366	17	AG065058
41	66.2	3.8	1086	17	AG077581
42	66.2	3.8	1101	17	AG039543
43	66	3.8	1164	17	AG128373
44	65.8	3.7	1339	17	BM458211
45	65.8	3.7	1452	17	AG032979

ALIGNMENTS

RESULT 1
LOCUS BG809984 1516 bp mRNA linear EST 22-MAY-2001
DEFINITION mgct002xd1f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
ACCESSION BG809984
VERSION BG809984.1 GI:14180964
KEYWORDS EST
SOURCE Magnaporthe grisea
ORGANISM Magnaporthe grisea
REFERENCE 1 (bases 1 to 1516)
AUTHORS Choi, W. and Dean, R. A.
TITLE Construction and sequence analysis of an appressorium stage cDNA library in the rice blast fungus, Magnaporthe grisea
JOURNAL Unpublished (2001)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: T3 primer (AATTAACCTCACTAAGGC).
Location/Qualifiers
1. 1516
/organism="Magnaporthe grisea"
/strain="70-15"

Query Match	4.78;	Score 83;	DB 12;	Length 1516;
Best Local Similarity	40.78;	Pred. No. 0.00033;		
Matches 530; Conservative	0;	Mismatches 765;		

[illegible]

111 GCGGCGCGCCGACCGTGGGCGGTGGCGGGGAGCCGATCGCCCGCATCGCGATCTGTGGA 170
194 GCGGGCGGCGCGCGCGGCG 253
171 CGCGCGCGCGGACACCGCGGTGCAAGTGTGCGGCGCTGTGTGTGCGCGCGCGCGCTCATGA 230
254 CG 313
231 CTCGCACACCCAGACGACACTACTGCTAGAGCGTGCACATGACGCCAAGATTC 290
314 NNN 373
291 GCAAGGCGGTACCAACGCGTGTGTACGCGGCAATTGGCGATC---AGCGTGGCGCGTGG 346
374 GCG 433
347 CGGACGCGCAACCG 406
434 CG 491
407 AGCGCTTGCAGCTACTACTGAGCGCGTGGCGGCGACGCGCGCGCGCGCGCGCGCGTCAACGCGCT 466
492 CGNN 551
467 GTATGTGGGCCATTCAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 526
552 GCGGGGCG 611
527 CCAGCGAGGAATCGGGGCGCATGCGGGAACCTGCGCGACGAGAAACCATGCGAGCGCGCA 586
612 CG 667
587 TCGGCAATTCGACGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 646
668 CG 727
647 TCGAGGTGTGCGCGCGCGTGAAGCGCATGCGCGCATCTACGCGCACCAACATCGCGAGC 706
728 CG 787
707 AAGCGAGACATCGGGGCGCGCGCGCTGAGAGAAAACCTTCGCGCATCGCGCGCGAGCTGAGC 766
788 CG 845
767 TCGCGGTGTGATCTGCGACACAAAGGTATGAGGCGGACGCCAATTCGCGCGCGTGTGCGCG 826
846 CCG 905
827 AGAGCGTGGCGTGAATCGAAGCG 905
906 CG 965
887 CCTACGTGAGCGCGTCAACATGCTCAAGCAGAGACCGCGTGTGCTGCGCGGAGCGACCA 946
966 NCG 1025
947 TCATCACTGTGTGAAGCGCTTCCCGAAG-----TGAGCGGCGCGCGACGCTGGAG 997

RESULT 3	AG171124/C		
LOCUS			
DEFINITION	AG171124	1798 bp	DNA, linear
	Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey		GSS 09-JAN-2002
ACCESSION	AG171124		
VERSION	AG171124.1		
KEYWORDS	GSS.		
SOURCE	Pan troglodytes male lymphocytes DNA, clone_1lb:RP43-Chimpanzee		
	Male BAC Library clone:RP43-040F09.TJ.		
ORGANISM	Pan troglodytes		
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.		
REFERENCE	1		
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,		
	Totoki, Y., Watanabe, H. and Sakaki, Y.		
TITLE	BAC end sequences of Library RPCI-43		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1798)		
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,		
	Totoki, Y., Watanabe, H. and Sakaki, Y.		
TITLE	Direct Submission		

RESULT 6	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
BG810038/c	BG810038	1116 bp MRNA	mgctc002x101f	1linear	EST 22-MAY-2001			
			grisea cDNA	clone mgctc002x101f 5', mRNA sequence.				
	BG810038							
	BG810038.1	GI:14181018						
	EST.							
	Magnaporthe grisea.							
	Magnaporthe grisea							
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes; Incertae sedis; Magnaporthaceae; Magnaporthe.							
	1 (bases 1 to 1116)							

[illegible]

QY 1337 CGGCCAGGCTGGCCGATACCCCACTTGAACACCTTACGAGGCGCCGCGGCATCC 1396
 Db 427 GCGCCNNCCCCCCCCCCCCNNCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNC 368
 QY 1397 ATTCGCTGATCAGTCAAGGCGCGCGCTGTGCAAGACAGCGCTTACCGGACGATG 1456
 Db 367 GCGCCCG 308
 QY 1457 CCGCGCGCTGCTGCAAGCGAGCGCGCGCTTACGAGGCGCGCGCGCGCGCGCGCG 1516
 Db 307 CCGGNCGCCG 248
 QY 1517 TGAACGGGGGCG 1576
 Db 247 CCC 188
 QY 1577 CTTCCGCTGCAATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1602
 Db 187 NNNNNNNCC 162

RESULT 7
 CNS0091P
 LOCUS
 DEFINITION CNS0091P 925 bp DNA linear GSS 03-JUN-1999
 Drosophila melanogaster genome survey sequence TET3 end of BAC #
 BACR19D16 of RPI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

ACCESSION AL053013.1 GI:4934461
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster.
 Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 925)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage:
 BP 191 91006 Evry cedex - FRANCE (E-mail: sequef@genoscope.cns.fr)
 Web: www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila genome project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Peter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 SOURCE
 1. 925
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR19D16"
 /clone_lib="RPI-98"
 /note="end: TET3"
 Location/Qualifiers
 120 a 61 c 172 t 511 others

BASE COUNT
 ORIGIN
 Query Match 4.1% Score 72.4; DB 17; Length 925;
 Best Local Similarity 13.3%; Pred. No. 0.017;
 Matches 51; Conservative 190; Mismatches 133; Indels 0; Gaps 0;

QY 1092 GCACGCGATCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1151
 Db 542 GGTGGTGTGTTTSSGSGYKGGCGSGBSGSCSSSCSSSSSCGCCGCCSCSSSTCC 601

QY 1152 GCGCCGCGATCCG 1211
 Db 602 SSBSSSKCSTSBSCSCCCSCSKSVCTGCTSCSSSSSSSSSSSSSSSSSSSSSSSS 661
 QY 1212 CCGGCGCTGTTCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1271
 Db 662 SSSSSSYTTRKTSASGSGSMAGGSGSTGTSSSSSSSSSSSSSSSSSSSSSSS 721
 QY 1272 CTTGGCGCTGCG 1331
 Db 722 SSS 781
 QY 1332 CGACCG 1391
 Db 782 GGTGSSSSSVGTSSSSSDSTGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 841
 QY 1392 CATCATCTCGGTGATGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1451
 Db 842 KCGCGCGGSSSTNGMGTSSACSSSSSSSSSSSSSSSSSSSSSSSSSSSS 901
 QY 1452 GCATGCG 1475
 Db 902 SASKSSSSG 925

RESULT 8
 AG039481
 LOCUS
 DEFINITION Pan troglodytes DNA, clone: PTB-016J15.F, genomic survey sequence.
 AG039481
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Pan troglodytes male lymphoblast DNA, clone: PTB-016J15.F.
 BAC library clone: PTB-016J15.F.
 Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
 1
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of library PTB
 2 (bases 1 to 1299)
 Unpublished
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimpbes@sc.riken.go.jp, URL: <http://bgp.gsc.riken.go.jp/>,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 clone tracking errors.
 PRIMERS
 Sequencing: -21M13
 LIBRARY
 Vector : pRS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 Location/Qualifiers
 1. 1299
 /organism="Pan troglodytes"
 /db_xref="taxon:9596"
 /clone="PTB-016J15.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
 24 a 571 c 592 g 31 t 81 others

FEATURES
 SOURCE
 1. 1299
 /organism="Pan troglodytes"
 /db_xref="taxon:9596"
 /clone="PTB-016J15.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
 24 a 571 c 592 g 31 t 81 others

BASE COUNT
 ORIGIN
 Query Match 4.1% Score 72.4; DB 17; Length 1299;

QY 1092 GCACGCGATCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1151
 Db 542 GGTGGTGTGTTTSSGSGYKGGCGSGBSGSCSSSCSSSSSCGCCGCCSCSSSTCC 601

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```

ACCESSION   AG030611
VERSION      AG030611.1
KEYWORDS     GI:16557484
SOURCE       Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
ORANISM      BAC Library clone:PTB-003A20.F.
REFERENCE    1
AUTHORS      Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
              Tokoki, Y., Watanabe, H., and Sakaki, Y.
TITLE        BAC end sequences of library PTB
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1341)
AUTHORS      Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
              Tokoki, Y., Watanabe, H., and Sakaki, Y.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
              1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.9sc.riken.go.jp/,
              Tel:81-45-503-9111, Fax:81-45-503-9170
              Clones are derived from the chimpanzee BAC library PTB This BAC end
              was generated during the R&D process and may have higher chance of
              clone tracking errors.
COMMENT      PRIMERS
              Sequencing: -21M13
LIBRARY      Vector : pKS145
              R.Site 1 : SacI
              R.Site 2 : SmaI
              Location/Qualifiers
              1. 1341
FEATURES     source
              /organism="Pan troglodytes"
              /db_xref="taxon:9598"
              /clone="PTB-003A20.F"
              /sex="male"
              /cell_type="lymphoblast"
              /clone_11b="PTB Chimpanzee Male BAC library"
BASE COUNT   48 a 476 c 700 g 14 t 103 others
ORIGIN
Query Match 4.1%; Score 71.8; DB 17; Length 1341;
Best Local Similarity 44.1%; Pred. No. 0.023;
Matches 245; Conservative 0; Mismatches 307; Indels 3; Gaps 1;
QY 1105 GCGTTCGCGCCGACATGATGCTCGACGAGCGCTGCGCCGACAGAGCGCCCGCATCCG 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1165 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1106
QY 1165 GCGCTGTGGGACCTTCCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1105 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1046
QY 1225 GCGGTGGAACGCGCGGTATGGAAGATGACGCGCTTACCGCGCGCTTGTGCGTGGCC 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1045 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 986
QY 1285 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 985 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 926
QY 1345 GTGGCGGATACCGCATCTTCAACACCTTACCGAGCGCGCGCGCGCGCGCATTCGCG 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 925 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 866
QY 1405 TACGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 865 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
QY 1462 GCGGTGCTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1521
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 805 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 746

```

```

QY 1522 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 745 NNGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 686
QY 1582 GCTCCGAATACGCGCCCAACCGCATTCGTGGCGCAAGAAGTATGGCGCGCGCGCGCG 1641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 685 CCGNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 626
QY 1642 GCGGAGCGCGCAAGGCC 1656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 625 CCGCGCGCGNCGCGCG 611
RESULT 11
LOCUS       CNS0091P/c
DEFINITION  Drosophila melanogaster genome survey sequence TERT end of BAC #
              BACR19D16 of RPI-98 library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
ACCESSION   AL053013
VERSION      AL053013.1
KEYWORDS     GI:4934461
ORANISM      Drosophila melanogaster.
              Drosophila melanogaster.
              Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 925)
AUTHORS      Genoscope.
              Direct Submission
              Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              Web : www.genoscope.cns.fr)
              Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutomo Osoegawa and
              Aaron Mamoser in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              p1 and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES     source
              Location/Qualifiers
              1. 925
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /clone="BACR19D16"
              /clone_11b="RPI-98"
              /note="end : TET3"
BASE COUNT   120 a 61 g 172 t 511 others
ORIGIN
Query Match 4.1%; Score 71.6; DB 17; Length 925;
Best Local Similarity 14.0%; Pred. No. 0.023;
Matches 55; Conservative 189; Mismatches 150; Indels 0; Gaps 0;
QY 1134 GCGCTGCGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 925 CBSBSCSCSBSBSCSSBSSBSSBSSBSSBSSBSSBSSBSSBSSBSSBSSBSSBSSBSS 866
QY 1194 GGGGCACTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 865 SSGTSACVCKNASSSCCGCGGABCCMCSSSSSCSASARGVKVAASGAGRGGS 806
QY 1254 GGGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	805	GGASASHSSSSSABSSSSSSCSASCHNASSSSSSASRSRHSGGAGGSGAASSRSSSSSS	746
Qy	1314	CGCCGACCTGTGTGTTCGACCGCCGACGCTGCGCCATCCGACCTTCGACACCC	1373
Db	745	ASAGSYVSSASSSSSSSSSSVSCSSVASSMSOSSSSSSASASSSSSSSSASCSGCC	686
Qy	1374	TACCGAGCGCGCGCGCGCATTCATTCCTGTACTCAACGCGCGCGCTGTGCGACA	1433
Db	685	TSMSCSTSTASASARSSSSSSSSSCSSSMASASSSSSSSSSSSSSSSSSSSSGSA	626
Qy	1434	GCAGGCTTCACCGCGCCACATCGCGCGCGCTGCTCGACGACGACGCGCGCTG	1493
Db	625	SMSSGGGSGSVASSMSSSSVSSSGRRSSSGGGGCGGSSGSSSSGSSGSSGSS	566
Qy	1494	GCACCAGCCCTTACATCCGCGCGTGAACGCGCGC	1527
Db	565	SGCMCRSCSSASAAAACVAAACGCMGCRKSG	532
RESULT 12			
AG076818			
LOCUS	AG076818	1152 bp	DNA
DEFINITION	Pan troglodytes DNA, clone: PTB-071C05.R, genomic survey sequence.		
ACCESSION	AG076818		
VERSION	AG076818.1	GI:16628620	
KEYWORDS	GSS.		
SOURCE	Pan troglodytes male lymphoblast DNA, clone.lib:PTB Chimpanzee Male Pan troglodytes		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.		
REFERENCE	1		
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.		
TITLE	BAC end sequences of library PTB		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1152)		
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
COMMENT	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.		
PRIMERS			
SEQUENCING	M13Rev		
LIBRARY	Vector : PKS145		
R SITE 1	: SacI		
R SITE 2	: SacI		
LOCATION/QUALIFIERS	1. 1152		
ORGANISM	"/organism="Pan troglodytes"		
DB_XREF	"/db_xref="taxon:9598"		
CLONE	"/clone="PTB-071C05.R"		
SEX	"/sex="male"		
CELL TYPE	"/cell_type="lymphoblast"		
CLONE LIB	"/clone_lib="PTB Chimpanzee Male BAC Library"		
BASE COUNT	34 a 488 c 514 g 23 t 93 others		
ORIGIN			
Query Match	4.1% Score 71.2; DB 17; Length 1152;		
Best Local Similarity	45.3% Pred. No. 0.028;		
Matches	308; Conservative 0; Mismatches 368; Indels 4; Gaps 3;		
Qy	105	CCGCGGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	164
Db	389	CG	448
COMMENT			

RESULT 15	Bg852363	LOCUS	DEFINITION	102852363	1281 bp	MRNA	linear	EST 29-MAY-2001
			Chlamydomonas reinhardtii CC-1690, normalized, lambda Zap II	Bg852363				
			Chlamydomonas reinhardtii CDNA, mRNA sequence.	Bg852363.1	GI:14233547			
			KEYWORDS					
			SOURCE					
				Chlamydomonas reinhardtii				

/strain="CC-1690 wild type mt+ 219r"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, lambda zap
 ii"
 /note="Vector: pbluescript II SK-, site 1: EcoRI, site 2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in YAP (acetate-containing) medium in the
 light, YAP medium in the dark, HS (minimal) medium in the
 ambient levels of CO₂ and HS medium bubbled with 5% CO₂.
 POLYA mRNA was purified from each sample, pooled and cDNA
 synthesized. The cDNA was directionally cloned into lambda
 zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pbluescript II SK- plasmids were excised from the lambda
 zap clones by superinfection with Exsist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al (1996) Genome Research 6: 791-806."
 697 C 456 g 15 t 83 others

Query Match	4.08;	Score 71;	DB 12;	Length 1281;
Best Local Similarity	42.88;	Pred. No. 0.031;		
Matches 515;	Conservative	0;	Mismatches 675;	Indels 14;
				Gaps 7;
QY	336	GGCGCCCTGGCGCAGCCCAACCGCGCGCCCTCTGAGACTGTCTGTGACGAAGCGGCTC	395	
Db	70	GGCGCGCGCGCCCGCCGNNCCGGGGCCCCCGCGCGCGCGCGCGCGCGCGCGCG	129	
QY	396	TTACCGTTTGAGCGGTGCGCCGACACTCACTGGACGCGTTGCGGGCCAGCGCGCGGCGT	455	
Db	130	CCCGCCCCCCCCCG	189	
QY	456	CAACGCGCGCTGTATGTGTGGGCAATCAACGCTGTGCGCGCGCGGTCATGCCGACTTGCA	515	
Db	190	CG	249	
QY	516	GGCGCGCGCGCACAGCAGAGAAATGCGGCGCATGGCGGCACTGGCGCGAGGAAGCATGGC	575	
Db	250	CAACGCGNNNGGAAAAAACMCACCCCTCTCTCTTTGTGCGNNNNNNNGGCGCGCNC	309	
QY	576	CAGCGCGCGCATCGGCATTTCAGCGCGCGCGCTTTCACCGCGCGCGCGCGCGCGCG	635	
Db	310	CANCCCCCCCCCGCGCGCNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	369	
QY	636	CGAAGAGATCATCAGGTGTGTCGCGCGCGCTGAGCGGCGCATGGCGCATCTTACGCCACA	695	
Db	370	GGCG	428	
QY	696	CATGCGCGAGCAAGCGCAGCATCCTGTGCGCGCGCTGTGAGAGAACTTCTCCGATCGCGG	755	
Db	429	CG	487	
QY	756	CGAGCTGGACGTCCCGGTGTGATCTCGCACCAAGTGCATGGGCGCACCCCAATTTCGG	815	

GenCore version 5.1.4.p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:03:15 ; Search time 753 Seconds
(without alignments)
11044.401 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 1758
Sequence: 1 gaattccactgacgcgcga.....ccctgacgcgcgaagctt 1758

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 6135869 seqs, 2365311053 residues

Total number of hits satisfying chosen parameters: 12271738

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New.*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq2:*
7: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq3:*
8: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq2:*
10: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	483	27.5	1425	US-10-282-122A-14027	Sequence 14027, A
2	425.4	24.2	1437	US-10-282-122A-24070	Sequence 24070, A
3	161.2	9.2	1590	US-10-156-761-3269	Sequence 3269, Ap
4	118.4	6.7	1728	US-10-156-761-3783	Sequence 3783, Ap
5	116	6.6	3354	PCT-US02-37547-12	Sequence 12, Appl
6	116	6.6	82746	PCT-US02-37547-56	Sequence 56, Appl
7	95.4	5.4	9975	PCT-US02-37547-4	Sequence 4, Appl
8	91	5.2	2538	US-10-156-761-6189	Sequence 6189, Ap
9	90.2	5.1	4770	PCT-US02-37547-3	Sequence 3, Appl
10	88.4	4.9	14055	PCT-US02-37547-1	Sequence 1, Appl
11	85.8	4.9	9222	US-10-329-079-12	Sequence 12, Appl
12	82.6	4.7	15738	US-10-329-079-6	Sequence 6, Appl
13	82.6	4.7	37360	US-10-156-761-6160	Sequence 6160, Ap
14	82.4	4.7	1371	US-10-398-605-7	Sequence 7, Appl
15	82	4.7	135638	US-10-398-605-3	Sequence 3, Appl
16	81.8	4.7	1248	US-10-398-605-50	Sequence 50, Appl
17	81.8	4.7	13613	US-10-271-889-50	Sequence 5, Appl
18	81.8	4.7	13613	US-10-271-889-46	Sequence 46, Appl
19	81.8	4.7	1512	US-10-156-761-4343	Sequence 4343, Ap
20	80.6	4.6	1464	US-10-156-761-1922	Sequence 1922, Ap
21	80.6	4.6	1290	US-10-282-122A-25795	Sequence 25795, A
22	80.4	4.6	1290	US-10-282-122A-25795	Sequence 25795, A

23	80.2	4.6	11817	US-10-156-761-2884	Sequence 2884, Ap
24	80.2	4.6	125746	US-10-156-761-15102	Sequence 15102, A
25	79.2	4.5	1458	US-10-419-128-1112	Sequence 1112, Ap
26	79.2	4.5	1458	US-10-366-683-1112	Sequence 972, App
27	79.2	4.5	1461	US-10-419-128-972	Sequence 972, App
28	79.2	4.5	1461	US-10-366-683-972	Sequence 930, App
29	79.2	4.5	4563	US-10-419-128-930	Sequence 4316, A
30	79.2	4.5	4563	US-10-366-683-930	Sequence 3773, Ap
31	79	4.5	2280	US-10-369-493-42316	Sequence 96, Appl
32	78.6	4.5	1404	US-10-156-761-3773	Sequence 74, Appl
33	78.2	4.4	12249	US-10-267-255-74	Sequence 42, Appl
34	78.2	4.4	18331	US-10-267-255-96	Sequence 34, Appl
35	78.2	4.4	18876	US-10-329-079-42	Sequence 2268, Ap
36	78.2	4.4	61944	US-10-329-079-34	Sequence 3184, Ap
37	77.8	4.4	3759	US-10-156-761-2268	Sequence 25489, A
38	77.8	4.4	5304	US-10-156-761-3184	Sequence 11, Appl
39	77.8	4.4	9025608	US-10-282-122A-25489	Sequence 21, Appl
40	77.4	4.4	4851	US-10-201-365-11	Sequence 4, Appl
41	76.2	4.3	5970	US-10-160-5398-21	Sequence 56, Appl
42	76.2	4.3	5970	PCT-US02-37547-56	Sequence 3854, Ap
43	75.8	4.3	9975	US-10-156-761-3854	
44	75.8	4.3	82746		
45	75.6	4.3	1719		

ALIGNMENTS

RESULT 1
US-10-282-122A-14027
Sequence 14027, Application US/10282122A
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Travick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14027
LENGTH: 1425

TYPE: DNA
ORGANISM: Burkholderia mallei
US-10-282-122A-14027

Query Match 27.5%; Score 483; DB 9; Length 1425;
Best Local Similarity 60.3%; Pred. No. 2,1e-66;
Matches 866; Conservative 0; Mismatches 520; Indels 51; Gaps 2;

40 CAATCGGATTCACAGCCCTTCACACCTGCTGCTCGGGGGGAGACCCCTCATCGAGGAGC 99
Db CACTCCGATCCCGAAGTGGCCGATGATGCTGCGGGGCGCAGATTGATGACAGGAGC 63
QY 100 AACACCCCGGG 159
Db GGG 123
QY 160 GATCTGTGAGAGCGGG 219
Db 124 AATCTGACGAACTGGCTTGGCGGAGAGGTGTGAGAGGCGAAGGGCGGGCGCTCGCC 183
QY 220 GGCCTTCATGACTCGACACCGACGAGCAACTACCTGCTCAGGCGCTCGGACATGACG 279
Db 184 GGCCTTCATGACTCGACACCGACGAGCAACTACCTGCTCAGGCGCTCGGACATGACG 243
QY 280 CCCAAGATCTGCGAGGGGCGTCAACACGGTGTCAAGCGCAATGCGGATGAGCGCTGGC 339
Db 244 CCGAAGATCTGCGAGGGGCGTCAACACGGTGTCAAGCGCAATGCGGATGAGCGCGCTG 303
QY 340 CCGGCGGCGACAGGCGCAACCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 399
Db 304 CCGGCGGCGACAGGCGCAACCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 329
QY 400 CGTTTCGAGCGCTTCGCGGACTACCTGAGACGCTTGGGGGGGGGGGGGGGGGGGGGG 459
Db 330 CGATTCGGA-----TGACCTGCTCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 372
QY 460 GCGGCGCTGATGCTGGGCGCATTAACGCTGCGGGCGCGGGTCAATGCGGACTTCAGCG 519
Db 373 GTCGGGGCGCTGCTGGGCGCATTAACGCTGCGGGCGCGGGTCAATGCGGACTTCAGCG 432
QY 520 GCGGCGCGACGAGAGAAATCGGGGGCGATGCGGAGCTGCGGCGAGAAAGCGATGCGC 579
Db 433 GCGGCGCGACGAGAGAAATCGGGGGCGATGCGGAGCTGCGGCGAGAAAGCGATGCGC 492
QY 580 GGGGCGCATGCGCATTTTCGACCGGGCGCTTTCACCGCGCGCGCGCGCAACGAGAA 639
Db 493 GGGGCGCATGCGCATTTTCGACCGGGCGCTTTCACCGCGCGCGCGCGCAACGAGAA 552
QY 640 GAGATCATGAGGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699
Db 553 GAGGTATGCGCGCTCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
QY 700 CCGGCGAGAGCGGAGACATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 759
Db 613 CCGGCGAGAGCGGAGACATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 672
QY 760 CTGAGGTCGCGGTGTGTATTCGACCAAGGTCATGAGCGCGCGCGCGCGCGCGCGCG 819
Db 673 GCACGGGTGCGGTGTGTATTCGACCAAGGTCATGAGCGCGCGCGCGCGCGCGCGCG 732
QY 820 TCGCGGAGAGCGCGCGGTGTATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 879
Db 733 AGCGCGAGAGCGCGGTGTATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 792
QY 880 GCGTATCCCTACGCGCGCGGTTCACCATGTCAGAGAGACCGCGGTCTGCTGGCGGGA 939
Db 793 TCGTATCCCTACGCGCGCGGTTCACCATGTCAGAGAGACCGCGGTCTGCTGGCGGGA 852
QY 940 CCGACCATCATCACTGCGAGAGCGCTTCCCGCACTGAGCGGGCGCGCGCGCGCGCGCG 999
Db 853 GACATTCAGATCACTGCTGCTGATCGATCCGCAATCGCGGAGAGCGCGCGCGCGCGCG 912
QY 1000 GTGCGGGCGCGAGCGCGCAATTCAGATGAGGTGTGGCGCGAGCTGCGACCGCGCGCG 1059

Db 913 ATGCGCGCGAGGTGGCGGGTGTCTCCAGAGAGCGGGCGCGCGCGCGCGCGCGCGCG 972
QY 1060 GCCATCTACTTCATGATGAGAGACCGGAGCGGAGCGGCGATCTGCGGCGCGCGCGCG 1119
Db 973 GCGGTGTATCAACAATCTCGGAGAGACAGCTGCGCGGATCTGTGCGACCGCGCGCGCG 1032
QY 1120 ATGATCGGCTCCGAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1179
Db 1033 ATGCGGCGCTCCGAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1092
QY 1180 TTCCCGGGGTGCTGGGGGACTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1239
Db 1093 TTTCGGCGCGGTGCTGCGGCTACTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1152
QY 1240 GTANGAAGATGACCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1299
Db 1153 GTGCGCAAGATGATGCGCGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1212
QY 1300 CAGGCGGGTACTTGGCG 1359
Db 1213 CCGGTGCGGTATCAAGCGGATCTGCTGCTGTGATGCGCGCGCGCGCGCGCGCGCGCG 1272
QY 1360 ACCTTCGACACCTTACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1419
Db 1273 ACCTTCGAGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1332
QY 1420 CCGGTGCGGAGAGAGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1476
Db 1333 CTGTCTTACCGCGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1389

RESULT 2

US-10-282-122A-24070
Sequence 24070, Application US/10282122A

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: EITRA.034A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

	Query Match	9.2%	Score 161.2	DB 8	Length 1590
	Best Local Similarity	50.3%	Pred. No. 2e-16		
	Matches 586	Conservative	0	Mismatches 358	Indels 42
				Gaps	6
QY	TCGACCCGCGCGCGCGCGCGCGACCCCTTCATGACGACGACACACCCCGAGCGCGCG				118
Db	2 TGGATCTCGTCATCTCGATGTCGCGCGTGTGAGGCGACCGGGGCGCTCTCGCCG				61
QY	119 CCGACCTCGGCGGTGCGGCGACCGCCATGCGCGCCATGCGCGATCGCGGATCGCGG				175
Db	62 CCGACGCTCGGTATCACCGCGCGCCCGATGTCGCGGATGTCGCGCGGAGCGCGCGCCCTGG				121

QY 176 CCGGACACACCGGCGGCTGACGCTGCGGCGGCGGCGGCTTCACTGACCTCC 235
 DB 122 GCGGACGCGGCGGCTGACGCGGCGGCGGCTGCGGCGGCGGCTTCACTGACCTCC 181
 QY 236 ACACCCAGACGACACCTGCTGCTGACGCGGCTGCGGCGGCGGCTTCACTGACCTCC 295
 DB 182 AGCGCCAGACGACCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 241
 QY 296 GCGTACACGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 355
 DB 242 GCGTACACCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 301
 QY 356 ACCGCGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 397
 DB 302 GCGTACGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 361
 QY 398 ACCGCTTACGACGCTTCCGCGGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 457
 DB 362 TCACCTGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 418
 QY 458 ACCGCGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 517
 DB 419 ACCGCGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 478
 QY 518 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 577
 DB 479 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 538
 QY 578 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 637
 DB 539 AGCGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 598
 QY 638 AAGGATCATGAGTGTGCGGCGGCGGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 697
 DB 599 CCGGACCTACGAGACTGTGCGGCGGCGGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 658
 QY 698 TCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 757
 DB 659 ACCGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 718
 QY 758 AGCTGACGCTGCGGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 817
 DB 719 AGCGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 778
 QY 818 GCTGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 874
 DB 779 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 838
 QY 875 TCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 934
 DB 839 TCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 889
 QY 935 CCGGACGACACCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 994
 DB 890 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 948
 QY 995 AAGGATCATGAGTGTGCGGCGGCGGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 1054
 DB 949 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 1003
 QY 1055 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 1114
 DB 1004 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 1063
 QY 1115 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 1174
 DB 1064 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 1123
 QY 1175 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 1200
 DB 1124 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 1149

RESULT 4
 US-10-156-761-3783
 ; Sequence 3783, Application US/10156761
 ; GENERAL INFORMATION:
 ; APPLICANT: OMOYA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 3783
 ; LENGTH: 1728
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermilt1115
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1728)
 US-10-156-761-3783
 Query Match 6.7%; Score 118.4; DB 8; Length 1728;
 Best Local Similarity 55.6%; Pred. No. 9e-10;
 Matches 345; Conservative 0; Mismatches 221; Indels 54; Gaps 4;
 QY 59 TCGACCTGCTGCTGCGGCGGCGGCGGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 118
 DB 5 TCGACCTGCTGCTGCGGCGGCGGCGGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 64
 QY 119 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 178
 DB 65 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 124
 QY 179 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 238
 DB 125 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 181
 QY 239 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 298
 DB 182 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 241
 QY 299 TCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 358
 DB 242 TCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 301
 QY 359 GCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 376
 DB 302 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 361
 QY 377 TCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 436
 DB 362 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 421
 QY 437 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 496
 DB 422 AGGCGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 478
 QY 497 CCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 550
 DB 479 AGGCGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 538
 QY 551 GCGGCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 610
 DB 539 TCGGCGGCTGCTGCTGACGCGGCGGCGGCTTCACTGACCTCC 598

b
2114 CGGCACCTGGGCCACCCACCCCTCGGCCACCCCGAGCTGCTACCCCTCTCT 2172

b
3191 CCTGGCGTCACCGCCTCACCAACGGCAGAT 3224

RESULT 6
PCT-US02-37547-56/C
Sequence 56, Application PC/TUS0237547
GENERAL INFORMATION:
APPLICANT: Floss, Heinz
APPLICANT: Yu, Tin-Mein
APPLICANT: Leister, Eckard
TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
FILE REFERENCE: DWASH-06712
CURRENT APPLICATION NUMBER: PCT/US02/37547
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn version 3.0
SEQ ID NO 56
LENGTH: 82746
TYPE: DNA
ORGANISM: Actinosynema pretiosum
PCT-US02-37547-56

Query Match Best Local Similarity 6.68; Score 116; DB 1; Length 82746;
Matches 788; Conservative 0; Mismatches 935; Indels 21; Gaps 6;

QY 56 CCTTCGACCTGCTGCTGCGGCGGACACCTCATCGACGGCAGCAACCCGGGGCGC 115
DB 73869 CCGACACCTGCTGACCGCGGGGCTGCCCGCCGCTGGCGCTGACGTCTGCTG 73810
QY 116 GCGCCGACCTGGGCGCTGCGGCGGACCGCATGCGCCGATGCGCATGCTCGAGCGCG 175
DB 73809 CCGCGCGGACGACGCGGCTGCGCGCGACGCTGCCACCGCGCGTGGCGCGCTGCGCG 73750
QY 176 CCGCGACACCGCGGCTGACGCTGCGGCGCTGCTGCTGCGCGCGCGCGCTTCATGATCGC 235
DB 73749 CCGCGACCGCGGCGGAGACGACGCGGACCAACCGCGCGCGCGCGCGCGCGCGCG 73690
QY 236 ACACCCACGACGACAACTGCTGCTGACGCTGCGACATGAGCGCCCAAGATTCGACG 295
DB 73689 ACCCGCGCGCTGCG 73630
QY 296 GCGTCAACAGGCTGCTGACGCGGCGGCAATGGCGGATGACGCTGCGCGCGCGCGCGCG 355
DB 73629 GCGCGACGCTGCG 73576
QY 356 ACCCG 415
DB 73575 TCG 73516
QY 416 CCGACCTGCTGACGCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 475
DB 73515 CCGACGCGCTGCG 73456
QY 476 GGCATTCAACGCTGCG 535
DB 73455 GCGACGCTGCTGCG 73396
QY 536 AATCGCGCGCATGCGGCG 595
DB 73395 CCGCGGAGACG 73336
QY 596 CGACCGGCGCGCTTACCG 655
DB 73335 GCGTACGCGCGCTGCG 73276
QY 656 GCGCGCGCGCTGACG 715
DB 73275 CCGCGACCTGCG 73216
QY 716 ACATGCG 772
DB 73215 TCGCGGAGCG 73156
QY 773 TGGTATCTCGACCAAGGTCATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 832

DB 73155 CCGAGCGCGCGGCACTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 73096
QY 833 TGCGCGTGAATGAGCG 892
DB 73095 AACCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 73036
QY 893 TGGCGCGCTGCG 952
DB 73035 GCG 72976
QY 953 CCGTGAATGAGCG 1009
DB 72975 CCGTGAATGAGCG 72916
QY 1010 AGCGCGCAATGAGCG 1069
DB 72915 GCG 72856
QY 1070 TCATGATGAGCG 1129
DB 72855 TGCGCGTGGCGCTGCTGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72796
QY 1130 CCGAGCGCGCTGCG 1183
DB 72795 GCGACGCTGACCG 72736
QY 1184 CCGCGCGTGTGCG 1243
DB 72735 TCGACG 72676
QY 1244 GGAAGATGACCG 1303
DB 72675 TCGTGGCGTGGCG 72616
QY 1304 CCGGCGTACTGCG 1363
DB 72615 GCGCGACG 72556
QY 1364 TCGAACACCGCTGACCG 1423
DB 72555 TGGTGTTC---GACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72499
QY 1424 TCGCGAAGAGCG 1483
DB 72498 CCGTGGCG 72439
QY 1484 CCGTGGCG 1543
DB 72438 CCGCGGACG 72379
QY 1544 CAACCGTGAAGCG 1603
DB 72378 TCG 72319
QY 1604 AATTCGTGGCG 1663
DB 72318 CCGTGGCG 72259
QY 1664 TGACCGTGAAGCG 1723
DB 72258 CCGCGGACCG 72199
QY 1724 AGCTGGCG 1757
DB 72198 CCGTGGCG 72165

RESULT 7
PCT-US02-37547-4
Sequence 4, Application PC/TUS0237547
GENERAL INFORMATION:
APPLICANT: Floss, Heinz

Db 1408 CCCGAGGACCTGCGCCCTGACCGCGAGCGGAGCGCCCGCCGACCCGATCACC 1467
 QY 388 GCGGCTCTTACCGTTGAGGCGCTTCCGCGACTCTGAGCGGCTTCCGGCCACCGCG 447
 Db 1468 GTGGGCTCTAGACCGGAGACTTACACAGAGGCGCCGCTACCCCTGCGCGCGCTC 1527
 QY 448 GCGGCGCTTACAGCGCGCTTATAGTGGGCCATTCAACGCGCGCGCGCGCTATGCGG 507
 Db 1528 GAATCGACATCCCGACAGACCCCGCCGACCAACGCCCTTGGCAAGCGCCCGCTG 1587
 QY 508 GACTTGCAGCGCGCGCGACAGAGAGAAATCGGCGCAATGCGGAGACTGGCGAGGAA 567
 Db 1588 CTCTCTCTCAACGAGCGGAGACTTACCTACGCGCAAGTCCGCTTGCACCCCGATCTTC 1647
 QY 568 GCCATGGCCAGCGCGCGCATATGCGCATTTGACCGCGCTTCTTACCGCGCGCGCGCG 626
 Db 1648 ACCACGCTCCGCGCGCGCGCTGCGCGCTTCCGACCGCCCTTACCGCGCGCGCTG 1707
 QY 627 -GCGCACACAGAGATCATGAGGTGTGCGCGCGCGCTGAGCGCGGATGCGCGCATC 684
 Db 1708 AAGCCCTTGGCGAGCGCGCGTACGAGCGCATCTCCCGCCACCGCTTACCTGACAGCC 1767
 QY 685 TACGCCACCCACATGCGCGAGAGAGGAGAGACATGCTGCGCGCGCTGAGGAAACCTTC 744
 Db 1768 GCG 1827
 QY 745 CGCATGGCGCGCGAGTGGAGTGGCGCGGTGTGATCTGCGACACAGAGTGTATGCGCG 804
 Db 1828 GCGACACCCAGTGGCG 1887
 QY 805 CCCAATTTGCG 1887
 Db 1888 ACCCTACAGACCTTGGCG 864
 QY 865 GACGCTGCTGAGCGCGGTATCTCTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1947
 Db 1948 CTGGCGCTGACCG 924
 QY 925 GTGCTGCG 2007
 Db 2008 AGCTGCTCTTCCGAGCG 984
 QY 985 GCGGAGCTGATGATGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2067
 Db 2068 GCGATCTGCG 1044
 QY 1045 CTGAGCG 2127
 Db 2128 CTGCTCCAGAGCCCGAGCG 1104
 QY 1105 GCGTGGCG 2187
 Db 2188 CCGGACCCCGAGTCCAGAGCG 1161
 QY 1162 CCGGCGCTTGGCG 2247
 Db 2248 AACTACTCTTACCG 1221
 QY 1222 TTCCGCGTGAACGCGCGTATGAGATGACCGCGCTGACCGCGCGCGCGCGCGCGCGCG 2307
 Db 2308 GCGGCGCTTACGTGAGCGCGTACTGACGAGCGCGGTGCGCGCGCGCGCGCGCGCGCGCG 1281
 QY 1282 GCGGCG 2367
 Db 2368 GCCATGCG 2414

RESULT 9
 PCT-US02-37547-3

; Sequence 3, Application PC/TUS0237547
 ; GENERAL INFORMATION:

; APPLICANT: Floss, Heinz

; APPLICANT: Yu, Tin-Wein
 ; APPLICANT: Leister, Eckard
 ; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor
 ; FILE REFERENCE: UASH-06712
 ; CURRENT APPLICATION NUMBER: PCT/US02/37547
 ; NUMBER OF SEQ ID NOS: 57
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO: 3
 ; LENGTH: 4770
 ; TYPE: DNA
 ; ORGANISM: Actinosynnema pretiosum
 ; PCT-US02-37547-3

Query Match 5.1% Score 90.2; DB 1; Length 4770;
 Best Local Similarity 42.9% Pred. No. 2.2e-05;
 Matches 562; Conservative 0; Mismatches 743; Indels 6; Gaps 2;

QY 111 GCGGCG 170
 Db 3204 GGGCGCGCTGCTGCGCGTGGAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3263
 QY 171 GCG 230
 Db 3264 CCG 3323
 QY 231 CTGCGACACCCAGAGACGACAACTACTGCTCAGCGCGTGGCGACATGAGCGCAAGATCTC 290
 Db 3324 CGAGGAGGTGAGGAGACAGGTGCGCATCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3383
 QY 291 GCAGGCGGTACACAGCGGTGCTGACGCGCGCAATTTGGCGCATGAGCTGCGCGCGCGCG 350
 Db 3384 GC---GCGCGAACCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3440
 QY 351 GCGCAACCG 410
 Db 3441 CACCGCGGAGACCG 3500
 QY 411 CTTCGCGACTACTGAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 470
 Db 3501 CGACGACCTGCTCTTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 530
 QY 471 GGTGGCCATTCAGCG 3560
 Db 3561 GCGGACCTGAGAGCG 530
 QY 531 CGAGGAAATCG 3620
 Db 3621 GCG 587
 QY 588 GCGCATTTGACCG 3680
 Db 3681 GGTGACG 647
 QY 648 CGAGGTGCG 3740
 Db 3741 GCGCGCGGTGCG 707
 QY 708 AGGAGACACATGCTGGCG 3800
 Db 3801 GCG 767
 QY 768 GCGGCGGTGATCTGCG 3860
 Db 3861 CCAGTGGCGCTTACG 3920
 QY 828 GACGCTGCGCGTGAATGAGCG 887
 Db 3921 GCAGGCGCGCGTGGCG 3980
 QY 888 CTACGTGGCGCGCGCTCACCATGCTCAAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 947


```

Db 3981 GCGACCCCGGACATCGCCCGCCAGCTCGCCAGCGCTGCGCTGCGCTGATGCCCCGGA 4040
Qy 948 CATCACTGTGTGCAAGCCCTTCCCGCAATGAGCGGGCCCGCACTGATGATGCGCGC 1007
Db 4041 CCGTGGCGGTGTCCGCGCTGTCTGACGCGCGCTGCGCCGCGAGAGACACCACCGCTGTGAC 4100
Qy 1008 CGAGCGCGGCAATCCAAATGACAGCTGTGCTGCGAGCTGACGCGCGCGCGGCGCATCTA 1067
Db 4101 CGACGTGCGTGGAGGGGTTCGCGCCCGGTTCACCGCCCTGCGCGCGAGCGCTGCT 4160
Qy 1068 CTTCATGATGAGCAACCGCAAGTGAAGCGCATCTGCGGTGTGCGCGCGCGAGCATATGCG 1127
Db 4161 CGCGACCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4220
Qy 1128 CTCCGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1187
Db 4221 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4280
Qy 1188 GGTGCTGGGCGCATCTAGCGCGCGCGCGCGCGCGCGCTGTTCGCGTGAAGCGCGGTATGGA 1247
Db 4281 GGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4340
Qy 1248 GATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1307
Db 4341 GGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4400
Qy 1308 GTACTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1367
Db 4401 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4460
Qy 1368 ACACCCCTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4511
Db 4461 CGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4511

```

```

RESULT 10
PCT-US02-37547-1
; Sequence 1, Application PC/TUS0237547
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wei
; APPLICANT: Leister, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; FILE REFERENCE: UMASH-06712
; CURRENT APPLICATION NUMBER: PCT/US02/37547
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 14055
; TYPE: DNA
; ORGANISM: Actinosynema pretiosum
PCT-US02-37547-1

```

```

Query Match 5.0%; Score 88.4; DB 1; Length 14055;
Best Local Similarity 44.4%; Pred. No. 4.1e-05;
Matches 460; Conservative 0; Mismatches 561; Indels 15; Gaps 2;
Qy 334 CGGATCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 383
Db 10878 CGACACCGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10937
Qy 384 CGAAGCGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443
Db 10938 CGACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10997
Qy 444 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 503
Db 10998 CGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11057
Qy 504 GCGCGACTTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 563

```

```

Db 11058 CGCCCTCATGCGCGAGCTGCCCCCGCGCGTGTGCGCGTCCGCGCGCGCGCGAGCA 11117
Qy 564 GGAAGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
Db 11118 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11177
Qy 624 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
Db 11178 CGTGTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11228
Qy 684 CTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743
Db 11229 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11288
Qy 744 CGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 803
Db 11289 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11348
Qy 804 GCGCAATTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 857
Db 11349 CACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11408
Qy 858 GCGCGAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 917
Db 11409 GGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11468
Qy 918 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 977
Db 11469 GGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11528
Qy 978 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1037
Db 11529 GCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11588
Qy 1038 GCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1097
Db 11589 GCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11648
Qy 1098 CATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1157
Db 11649 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11708
Qy 1158 GATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1217
Db 11709 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11768
Qy 1218 CCGTGTCCCGCTGAGAGCGCGGTATGGAATGACCGCGCGCGCGCGCGCGCGCG 1277
Db 11769 GATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11828
Qy 1278 CCGTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1337
Db 11829 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11888
Qy 1338 GCGCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1353
Db 11889 GGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11904

```

```

RESULT 11
PCT-US02-37547-2
; Sequence 2, Application PC/TUS0237547
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wei
; APPLICANT: Leister, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; FILE REFERENCE: UMASH-06712
; CURRENT APPLICATION NUMBER: PCT/US02/37547
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0

```

SEQ ID NO 2
 LENGTH: 9222
 TYPE: DNA
 ORGANISM: Actinosynema pretiosum
 PCT-US02-37547-2

Query Match 4.9%; Score 85.8; DB 1; Length 9222;
 Best Local Similarity 43.8%; Pred. No. 0.0001;
 Matches 614; Conservative 0; Mismatches 772; Indels 15; Gaps 5;

QY 328 ATCAACCTGGGCGCGCTGCGCCAGCCCAACCCGCGCCCGCTGAGACTGTGAGAA 387
 Db 7045 ACCGTCCACAGCCCAAGCCCGCTGTGCGCTTGGAGGTCGGGCGGCTCTCTGAG 7104
 QY 388 GCGGCGCTTACCTTTCGAGCGCTTTCGAGCTGAGCGGCTTTCGAGCGCGC 447
 Db 7105 GGGTGGGCGCTGTCGCGGAGCTGTCTGCGGCTGCGGCTGCGGCGCGCGC 7164
 QY 448 GCGGCGCTTACAGCGCGCTGTATGTGGGCTTTCAGAGCTGCGCGCGGCTGATGCC 507
 Db 7165 CACGTGCGGCGCGCTGTGTGAGGCGGAGCGCTTGGCGCTGCTGCGCGCGCGCG 7224
 QY 508 GACTTGCAGCGCGCGCGCCAGAGGAAATGCGCGCCATGCGGGAAGCTGCGCGAGAA 567
 Db 7225 GCCATGACAGGCGCGCGCGCGCGCGCGCGCGCTGCGGCTGCTGCGCGCGCGCG 7284
 QY 568 GCCATGCG 627
 Db 7285 GTGCGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7344
 QY 628 GCCACCAAGAGAGATCATAGGTGTGCC---GCCGTGAGCGCGCGCGCGCGCGCG 683
 Db 7345 GTGCTGTGCGGTGACGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7404
 QY 684 CTACG--CCACCAAGAGAGATCATAGGTGTGCC---GCCGTGAGCGCGCGCGCGCG 741
 Db 7405 TGGCGGACCAAGGCGCGCTGTGCGGTGAGCCAGCTTTCACCTCCGCGCGCGCGCG 7464
 QY 742 TTCGCCATGCG 801
 Db 7465 CTCGGTGAAGTGGCG 7534
 QY 802 CAGCCCAATTCG 861
 Db 7535 CTGTCCAGCTACCG 7584
 QY 862 CAGAGCTGTGCG 921
 Db 7585 GCTGCGAGCGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7644
 QY 922 CGCGTGTGCG 981
 Db 7645 GCGCTGACGCGG---GTGCTGAGAGTGGGCGCGCGCGCGCGCGCGCGCGCGCG 7701
 QY 982 GCGCGCACTGATGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1041
 Db 7702 GAGAGAGTCCG 7761
 QY 1042 GACTGACG 1101
 Db 7762 GGGCTGTCTCCGCGAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7821
 QY 1102 CTGCGCTTGGCG 1159
 Db 7822 CTGTGTGGGCG 7881
 QY 1160 ATCCGCGCTGTGGGCGCACTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1219
 Db 7882 TTCTGGCCCG 7941
 QY 1220 TGTTCGCGT---GAGACGCGCGGTATGAGATGACCGCGCGCGCGCGCGCGCGCG 1275
 Db 7942 GCGCGCTGGGCG 8001

QY 1276 GCGCTGGCG 1335
 Db 8002 GAGCTGGCGAGCGCTGCGCGAGCTGTCTGCGCGAGACAGCGCGCGCGCGCGCGCG 8061
 QY 1336 CCGGCGAGCG 1395
 Db 8062 GGGAGCG 8121
 QY 1396 CATTCTGTATGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1455
 Db 8122 GCGGAGGTGACAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8181
 QY 1456 GCGGCG 1515
 Db 8182 GAGCCGAGGT 8241
 QY 1516 GTGAGCG 1575
 Db 8242 GCGACCG 8301
 QY 1576 CCCTCGCTGCGAATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8361
 Db 8302 GTGCG 1695
 QY 1636 CTGCG 1695
 Db 8362 CTGTGTGCG 8421
 QY 1696 ATCGCGGTCTGACCGCTGTCC 1716
 Db 8422 GACGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8442

RESULT 12

US-10-329-079-12

Sequence 12, Application US/10329079

GENERAL INFORMATION:

APPLICANT: FARNET, Chris

APPLICANT: ZAZOROULOS, Emmanuel

APPLICANT: STAFRA, Alfredo

TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEP

FILE REFERENCE: 3002-1105

CURRENT APPLICATION NUMBER: US/10/329, 079

NUMBER OF SEQ ID NOS: 66

SOFTWARE: PatentIn version 3.0

SEQ ID NO 12

LENGTH: 15738

ORGANISM: Streptomyces fradiae

US-10-329-079-12

Query Match 4.7%; Score 82.6; DB 9; Length 15738;
 Best Local Similarity 43.1%; Pred. No. 0.00033;
 Matches 580; Conservative 0; Mismatches 749; Indels 18; Gaps 3;

QY 55 CCCTTGACCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 114
 Db 1174 CCGGAGCG 1233
 QY 115 GCGCGGAGCTGCG 174
 Db 1234 GCGCGGAGCTGCG 1293
 QY 125 GCGCGGAGCG 234
 Db 1294 GCGCGGAGCG 1353
 QY 235 CACACCGAGCG 294
 Db 1354 CTGAGGAGGTTGCG 1413

QY 295 GGGTCACACGAGGTGTACAGGCGCATTTGGGCAATCAACCTTGGGCGCCGCTGGCCACGCC 354
 Db 1414 GAGAGAGGGGTGTGAGAGGCGGTGTGTGACACCTTGGCCCGGAGGGGGCCACCCCGGAG 1473
 QY 355 AACCCGCGCCCGCCCTGAGACCTGTGTGAGAGGAGGCGGTCTTACGCTTTGAGAGCGCTC 414
 Db 1474 GCGCCGCGCGCTGT 1533
 QY 415 GCC-----GACTACCTGTGAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 468
 Db 1534 TCCCGCGCGCGCGGT 1593
 QY 469 ATGTGGGCGCATTTCAACGCTGT 528
 Db 1594 TGGCTGT 1653
 QY 529 GAGAGAGAAATCCGCGCATGT 588
 Db 1654 GTTCACAGTACCGGT 1713
 QY 589 GGCATTTCGACCGCGGCTTCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648
 Db 1714 GTTGGCG 1773
 QY 649 GAGGT 708
 Db 1774 GCGCGCGGTCTGT 1833
 QY 709 GCGGAGACATGT 768
 Db 1834 GCGCGGT 1893
 QY 769 -----CGGT 819
 Db 1894 ATCCACACCTGT 1953
 QY 820 TCGCGGAGAGCGT 879
 Db 1954 TCGCGGT 2013
 QY 880 GCGTATCCCTAGT 939
 Db 2014 GCG 2073
 QY 940 GCGACCATCATCACTGT 999
 Db 2074 CTGCGATGT 2130
 QY 1000 GTGCG 1059
 Db 2131 CCG 2190
 QY 1060 GCGATCTACTCATGT 1119
 Db 2191 ACCCTACTTCCGACACCTGT 2250
 QY 1120 ATGATGT 1179
 Db 2251 CTGCTGT 2310
 QY 1180 TTCCCG 1239
 Db 2311 GCG 2370
 QY 1240 GTATGT 1299
 Db 2371 GTGACGCTACTGT 2430
 QY 1300 CAGCGCGGT 1359
 Db 2431 GCGT 2490
 QY 1360 ACCTTGAACACCTTACCGAGCGCGC 1386

Db 2491 CTGCGCGGACCGCGAGT 2517
 RESULT 13
 US-10-329-079-6
 ; Sequence 6, Application US/10329079
 ; GENERAL INFORMATION:
 ; APPLICANT: FARNER, Chris
 ; APPLICANT: ZAZOPOULOS, Emmanuel
 ; APPLICANT: STAFPA, Alfredo
 ; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTI
 ; FILE REFERENCE: 3002-1105
 ; CURRENT APPLICATION NUMBER: US/10/329,079
 ; NUMBER OF SEQ ID NOS: 66
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 37360
 ; TYPE: DNA
 ; ORGANISM: Streptomyces fradiae
 US-10-329-079-6

Query Match 4.7%; Score 82.6; DB 9; Length 37360;
 Best Local Similarity 43.1%; Pred. No. 0.00033;
 Matches 580; Conservative 0; Mismatches 743; Indels 18; Gaps 3;
 QY 55 CCTTGTGACCTGT 114
 Db 15390 CCGAGCG 15449
 QY 115 CCGCGGACCTGT 174
 Db 15450 GCGCGGACCTGT 174
 QY 175 GCGCGGACCTGT 234
 Db 15510 GCGCGGACCTGT 234
 QY 235 CACACCCAGAGACATCTGT 294
 Db 15570 CTGAGAGAGT 15629
 QY 295 GCGGTACACAGGT 354
 Db 15630 GACGAGGGGT 15689
 QY 355 AACCG 414
 Db 15690 GCG 15749
 QY 415 GCC-----GACTACCTGTGAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 468
 Db 15750 TCCCG 15809
 QY 469 ATGT 528
 Db 15810 TGGCTGT 15869
 QY 529 GAGAGAGAAATCCGCGCATGT 588
 Db 15870 GTTCACATGT 15929
 QY 589 GGCATTTCGACCGCGGCTTCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648
 Db 15930 GTGCG 15989
 QY 649 GAGGT 708
 Db 15990 GCGCGGT 16049
 QY 709 GCGGAGACATGT 768

```

Db 16050 GGGCCGGTACCGGACCGGACCGCCCTCTCTGCGGGGACGCGCCCTAGTCTC 16109
QY 769 -----CCGGTGGTATCTCGACACCAAGGTCATGGGCGACGCCAATTTCGGCCG 819
Db 16110 ATCCACACCTGGGTTTCCACCGGCGGACGCGCCCAAGGGGTCACGGTGGACACCGGCGCTG 16169
QY 820 TCGCGCGAGAGCGTCCGCTGATGAGCGCCCATGCGCGCGACGAGAGTCTGCTGAGC 879
Db 16170 TCGCGGCTGCTCCAGGCGCCACCGCGGGTCACTTCCTCCGCACTCCCTCCGACGGC 16229
QY 880 GCGTATCCCTACGTGGCGGGGTCCACCATGCTCAAGCAGAGACCGGCTGCTGGCGCGCA 939
Db 16230 GGGCCCGGCGCGCGCGCCCGCCACGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16289
QY 940 CGCACATCATCATCTGATGAGCGCCCTTCGCCGAACTGAGCGCGCGCGCGCGCGATGAA 999
Db 16390 CTCGCGATGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16346
QY 1000 GTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1059
Db 16347 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16406
QY 1060 GCGATCTACTTCTCATGTGAGCAGAACCGCGAGCGAGCGAGCGAGCGAGCGAGCG 1119
Db 16407 ACCTACTTCCGACCGCTGCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16466
QY 1120 ATGATGGGCTCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1179
Db 16467 CTGCTTGGCCCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16536
QY 1180 TTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1239
Db 16527 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16586
QY 1240 GTATGAGATATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1299
Db 16587 GTGACCGTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16646
QY 1300 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1359
Db 16647 GCGTACGTCCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16706
QY 1360 ACCTTCAACACCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386
Db 16707 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16733

```

RESULT 14

US-10-156-761-6160

Sequence 6160, Application US/10156761

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

PRIOR FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 6160

LENGTH: 1371

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1371)
US-10-156-761-6160

Query Match 4.7%; Score 82.4; DB 8; Length 1371;
Best Local Similarity 46.4%; Pred. 0.00035;
Matches 309; Conservative 0; Mismatches 351; Indels 6; Gaps 1;

```

QY 66 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125
Db 687 GCGCATGCGGTATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 746
QY 126 GGGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 185
Db 747 CAACCGGTTCGCGAGTGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
QY 186 CCGGCTGACGCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 245
Db 807 CCGGAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 866
QY 246 CGACACTACTGCTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 305
Db 867 CGCGACCGCGCTCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 924
QY 306 GGTGCTCACGGGCAATTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 365
Db 925 ----GGCCACCGACGCGATCCACACACTGCTGCGCGCGCGCGCGCGCGCGCGCG 980
QY 366 CCGCGTGAAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425
Db 981 CTCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040
QY 426 GGACGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 485
Db 1041 CGACACGCTGTGACGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1100
QY 486 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 545
Db 1101 CCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1160
QY 546 CATGCGGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 605
Db 1161 CGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1220
QY 606 CTCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665
Db 1221 CATGCGCGGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1280
QY 666 GAGCGCGCATGGGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 725
Db 1281 CAAGAACCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
QY 726 CCGCGCT 731
Db 1341 CCACCT 1346

```

RESULT 15

US-10-314-657-1/C

Sequence 1, Application US/10314657

GENERAL INFORMATION:

APPLICANT: SHEN, Ben

APPLICANT: CHENG, YI-Qiang

APPLICANT: TANG, Gong-qi

TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide

FILE REFERENCE: 054030-0021

CURRENT APPLICATION NUMBER: US/10/314,657

PRIOR FILING DATE: 2002-12-09

PRIOR APPLICATION NUMBER: PCT/US02/08937

PRIOR FILING DATE: 2002-03-22

PRIOR APPLICATION NUMBER: US 60/278,935

PRIOR FILING DATE: 2001-03-26

NUMBER OF SEQ ID NOS: 214

```

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

```

Query Match	4.7%	Score 82;	DB 8;	Length 135638;
Best Local Similarity	44.3%	Pred. No. 0.00041;		
Matches 610;	Conservative	0;	Mismatches 750;	Indels 18;
			Gaps	6

	Conservative	U/ mismatches	750; Indels	18; Gaps
QY 117	CGCCGACCTGGGGGTGGCGGCGACCCGATCCCGCATCCGGCATCGGCGATCTGGAGCGCCG	176		
Db 118012	CGACCACTCCGCTCCCAAGCGGCACTCCACCCGGACCGCACTGTCATCTGAC	117955		
QY 177	CGGCGACACCCGGGGTGGAGCTGTGGGGCTGGTGGTGGGGCCGGCTTTCATCGACGCA	236		
Db 117952	CTGCCCGACCTCTCTGTATGGGAGAGGGAGACCGGGCGCTGGAGGCACTGGCCCGGGGCT	117893		
QY 237	CACCCAGAGACACTACTCTGTAGAGGGTGGCGACATGAGGCCAAGATCTGGCAGG	296		
Db 117892	GCTCGGGCGCGGACAAAGCCGGGGGACCCCGTATCGGCAATCTTCTTGGTGGAGAGCTG	117833		
QY 297	CGTACCAACGGTGGTACGGGCAATTGGGGCATCAGCTGGCGCGCTGGGGCGACGGCAA	356		
Db 117832	GGCGGGCTACGGGGTGGCTTACTCTGGGGCGCTCAAGGGCGGGCGAACCGTCTTACCT	117773		
QY 357	CCCGCCGGCCCGCCCGAGACCTGTGGAGAGAGGGGGCTTTCACGTTTCGAGCGCTTGC	416		
Db 117772	GGCGCTGGCGCTCCCGCGGACGACACTGAAAGGGGTGGCCCTCGAGAGCGAGC---	117716		
QY 417	CGACTACCTGAGCGGTTTGGGGGCCACCGCGGGCGGTCAACCGCGCTGTATTGGGGG	476		
Db 117715	CGGCAATCGTCCACGGGGGAGACCGCCCGCCACGACCTCGGGCGTCCCTCTGACCGGGAC	117656		
QY 477	CGATTCAACGGTGGCGCGGGGGGGGTATGCGGACTTCAGGGCGCCGACGACGAG	536		
Db 117655	GCTCGACGAGCTGTCCGGCCCGGGGGAGACACCGGTGAGACTCGTCCACAGCCCGGGG	117596		
QY 537	AATCGCGGCATGTGGGGAGCTGTGGCGAGAGAGCCATGGCCAGGGGGCGCATCGGCACTTTC	596		
Db 117595	CGCGCGGAGATCGTGTACTCTCTCGGACACACCGAGCTGGCCCGGGCGCTCTCTCTC	117536		
QY 597	GACCGCGCCCTTTCACCGCCCGCGCGCGCGCCGCCACCGACGAGATATCAGAGTG	656		
Db 117535	GCACCAAGAACTCGGAGCGGGGGCGGACCGCCCTCGGTATGGCCACGACGAGGCCAC	117476		
QY 657	CC---GGCGGTGAGCGGGCATGGCGGCATCTAGCCACCCACATGCGCGAGAGGCGA	713		
Db 117475	CCCATGTGTGGCTCGGGAACCTGGGCACTACCGGAGGCGCACACCGTCCATGTT	117416		
QY 714	GCACATCGTGGCCCGGTGGAGGAAACCTTCCGCATGCGCGGGAGACTGGAGTCCGCT	773		
Db 117415	GCTCAAGCCACCCCGACACAGCTCGTCTCGCGCCCGGGGAGAGCCGACCGGCTGTG	117356		
QY 774	GGTATCTCGACACCAAGGTCAATGGGCGACCCCAATTTGGCGCGCTCGCGGAGAGCT	833		
Db 117355	CGCGCTATCGAGACCATCGCGCGAGACCGGCATGATGAGACCCGAACTCTCGCGTGA	117296		
QY 834	GGCCCTATCGAGGGCGCCATGGCGGGCCAGGACGTCCTGTGGAGCGGTATCCCTACGT	893		
Db 117295	GATACCCGCGAGCGCGCTCGCGCGTACGACCTGAGAGTCCGCTAC---CACGCTGGC	117239		
QY 894	GGCGGCTCAGCATGTCTAAGCAGAGACGCGGTCGTGGCGGAGACGACCATCATAC	953		
Db 117238	CACCGCTCGGGCTTCCTCATCCCGCGTCCCGCGCGCTCTTGGCGCGGATGGCCCG	117179		
QY 954	CTGTGTCAACCCCTTCCCGGAACTAGCGGGCGCGACCTGATGTAAGTGGGGCGAGCG	1013		
Db 117178	GGCCCGGTATGGGGCGCTTACTCGGCGACCGAGCGCAAGCCCGCGTGAACATCGGAC	117119		
QY 1014	CGGCAATCAAGATAGAGGTGTGGCCGAGTGAACCGCGCGCGCATCTACTCAT	1073		

Db 117118 CTTGCAACCCGGGGCCCCGGATATGTGCGCCGGCCGCCGCCGCCGCCCGGACACCATATGACTGAT 1170558
QY 1074 GATGAGCAACCCGCAAGTg---CAAGCATTCCTGGAGCTTGGCCCGGACATGATCGGGCTC 1130
Db 117058 CACCGGACGACAGCGCGCGGGAGACTGCCGCCCATCGCTGGCTGCGGCGGATCTGGCTGCCGC 1169999
QY 1131 CGACGGCTTCCCGCAGCAGAGAGCGCCGATCCGCGCTGTGGGGACAC---TTCCCGCG 1187
Db 116998 CGACGGCGCGCTCCCGCCGACACCGCGCTGAGAGCGGGGCCGAGGACCAACCGTGTCCGGGA 1169399
QY 1188 GGTGCTGGGGGACTATGCGCCCGACCTGGCGCTTCCCGCTGGAGACGGCGGTATGGAA 1247
Db 116938 GGGCGGATGTGTGCGACACCGCGCGACCTCGGACAGTGAAGAGAGGCGGACGTATATC 116879
QY 1248 GATGACGGGCGCTGACCGCGCGCGGCGCTTGGCGCTGGCGCGCGCGGCGAGCTGACGGCCGG 1307
Db 116878 CTTGCACCGCGGAGACCGGACGGCGCGTGCACACACCGCGCGGCGCTGTGTCGTCCTG---CG 116822
QY 1308 GTACTTCCGCAAGCTGGGAGGTGTGACACCGCGGCGACAGGTGGCCGATACCGCCACTTTCGA 1367
Db 116821 GGTGAGTCGCGTCTCTGCTCGACGACGAGGGCGGTGCGCGAGCGGCGCGGTGTCTGCCCGCCG 116762
QY 1368 ACACCCATACGAGCGCGCGCGCGGCACTCATTCGCTAGCTGACGAGGGCGCGCGGTCTG 1427
Db 116761 GCGCGCGCGCGGTGGCGCGCGCGCGGATCGTGGCCCGCGCGCGCGGCCACCACTGACCCGGAAGCT 116702
QY 1428 GCAAGAGCAGGCGCTTCAACCGCGGCAAGCATGGCGCGCGGTCTCTGCACGACGAGGCGCAC 1485
Db 116701 GCTGGCGGCGACCTCTCGCGCGCTGACGCGGAAGACACGCTGCGCCCGCCACGAGATGCC 116644

Search completed: May 11, 2003, 12:05:20
Job time : 935 secs

Result No.					Score		Query Match		Length		DB		ID		Description
1	1758	100.0	1758	38	US-10-009-782-1	Sequence 1, Appl									
2	1440.6	81.9	1473	31	US-09-807-788-3	Sequence 3, Appl									
3	428.6	24.4	1479	18	US-09-489-039A-1847	Sequence 1847, Appl									
4	133.6	7.6	453	62	US-60-184-777-35	Sequence 35, Appl									
5	88.8	5.1	5871	41	US-10-152-886-54	Sequence 54, Appl									
6	86	4.9	88421	36	US-09-927-059-1	Sequence 1, Appl									
7	85.4	4.9	770	29	US-09-728-449-925	Sequence 925, Appl									
8	85.4	4.9	770	31	US-09-803-110-925	Sequence 925, Appl									
9	84.6	4.8	109319	29	US-09-758-759-1	Sequence 1, Appl									
10	84	4.8	1839	29	US-09-758-759-150	Sequence 150, Appl									
C 11	82	4.7	135638	29	PCR-US02-08937-1	Sequence 1, Appl									
12	81.8	4.7	1248	32	US-09-836-821-7	Sequence 7, Appl									
13	81.8	4.7	1248	33	US-09-860-846-7	Sequence 7, Appl									
14	81.8	4.7	1248	33	US-09-861-289-7	Sequence 7, Appl									
15	81.8	4.7	1248	37	US-09-988-848-7	Sequence 7, Appl									
C 16	81.8	4.7	12441	37	US-09-988-848-3	Sequence 3, Appl									
C 17	81.8	4.7	13613	32	US-09-836-821-3	Sequence 3, Appl									
C 18	81.8	4.7	13613	33	US-09-860-846-3	Sequence 3, Appl									
C 19	81.8	4.7	13613	33	US-09-861-289-3	Sequence 3, Appl									
20	80.2	4.6	5760	41	US-10-152-886-14	Sequence 14, Appl									
21	80	4.6	3240	33	US-09-887-372A-4417	Sequence 4417, Appl									

Sequence 1112, April
Sequence 972, April
Sequence 126, April
Sequence 45, April
Sequence 930, April
Sequence 1, April
Sequence 42316, April
Sequence 84, April
Sequence 7458, April
Sequence 721, April
Sequence 193, April
Sequence 193, April
Sequence 74, April
Sequence 74, April
Sequence 74, April
Sequence 96, April
Sequence 96, April
Sequence 96, April
Sequence 1, April
Sequence 2, April
Sequence 8, April
Sequence 8, April
Sequence 8, April

RESULT 1

```

: Sequence 1, Application US/10009782
: GENERAL INFORMATION:
: APPLICANT: TAKEUCHI, Ken-ichi
: APPLICANT: KOIDE, Yoshino
: APPLICANT: HIROSE, Yoshihiko
: APPLICANT: MORISUCHI, Mitsunaki
: APPLICANT: ISOBE, Kimiyasu
: TITLE OF INVENTION: TRANSFORMED MICROORGANISM AND PROCESS FOR PRODUCING D-AMINOACYLASE
: FILE REFERENCE: 217301USOPCT
: CURRENT APPLICATION NUMBER: US/10/009,782
: PRIOR FILING DATE: 2001-12-17
: PRIOR APPLICATION NUMBER: PCT/JP00/03932
: PRIOR FILING DATE: 2000-06-15
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 1758
: TYPE: DNA
: ORGANISM: Alcaligenes xylosoxydans subsp. xylosoxydans
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (34)..(1485)
: OTHER INFORMATION:
US-10-009-782-1

```

Query Match	100.0%;	Score 1758;	DB 38;	Length 1758;
Best Local Similarity	100.0%;	Pred. No. 4.1e-259;		
Matches 1758; Conservative	0;	Mismatches 0;		

	Seq. coordinates	U	Indels	Gaps
QY	1 GAATTCACCTTGATGCGCGAAGAGAGATTTCCATGTCCCAATCGATTCACGAGCCTTC			60
Db	1 GAATTCACCTTGATGCGGAGAGAGATTTCCATGTCCCAATCGGATTCACGAGCCTTC			60
QY	61 GACCTGCTGCTCGCGGGGGGACCCCTCATTCAGCGGACACACCCCGGGGGCGCGCC			60
Db	61 GACCTGCTGCTCGGGGGGGACCCCTCATTCAGCGGACACACCCCGGGGGCGCGCC			120
QY	121 GACCTGGGCGTGC CGCGGAGCCGCATCGCGCGGCATCGGCGATCTGTGGAGCGCGCGG			180
Db	121 GACCTGGGCGTGC CGCGGAGCCGCATCGCGCGGCATCGGCGATCTGTGGAGCGCGCGG			180
QY	181 CACACCGGGGTCGACGTGTGCGGGGCTGTGTGTCCGCGCGGCTTCATCGATCCGACACC			240

[illegible]

Db	1261	ACGGCCGGCGCTTCGGCCCTGCGCCGGCGCGGCGAGCTGCAGGCGGGGTACTTCCTGCCGAC	1320
Qy	1321	CTGGTGGTGTTCGACCCGGCCACGGGTGGCCGATACCGCCACTTCGAAACACCTACCGAG	1380
Db	1321	CTGGTGGTGTTCGACCCGGCCACGGGTGGCCGATACCGCCACTTCGAAACACCTACCGAG	1380
Qy	1381	CGGCGCGCGGGCAGTCCATTCCTCCGTGTAGTCAAGCGCGCGCCGGTCTGGCAAGACGAGCG	1440
Db	1381	CGGCGCGCGGGCAGTCCATTCCTCCGTGTAGTCAAGCGCGCGCCGGTCTGGCAAGACGAGCG	1440
Qy	1441	TTACCGCGCGCAGCATGCGCGCGCGCTGCTGCACAGCAGCGCCGGCTGAGGCGCGGCGCAG	1500
Db	1441	TTACCGCGCGCAGCATGCGCGCGCGCTGCTGCACAGCAGCGCGCCGGCTGAGGCGCGGCGCAG	1500
Qy	1501	CCCTTCAAAATCGGCGCGTGAAAGGGGGGGCGCTGCCGCCCTCTCCCAACCTTGAGCGCAAC	1560
Db	1501	CCCTTCAAAATCGGCGCGTGAAAGGGGGGGCGCTGCCGCCCTCTCCCAACCTTGAGCGCAAC	1560
Qy	1561	CGGTACATGAGCCCTCCCTCCGCTCGCATACGAGGCCACCGATATCGTGGGCAAGGAA	1620
Db	1561	CGGTACATGAGCCCTCCCTCCGCTCGCATACGAGGCCACCGATATCGTGGGCAAGGAA	1620
Qy	1621	GGATGGGGGGCGGCGCTGGGGCGCGGAGCGCAGGCCCGGAAAAATGACCCCTGCAAGACTG	1680
Db	1621	GGATGGGGGGCGGCGCTGGGGCGCGGAGCGCAGGCCCGGAAAAATGACCCCTGCAAGACTG	1680
Qy	1681	TGCGAAGCCAGAGCGGATCGCGGCTGTGACCCCTGTCCAAAGGCCAGTGTGGGCCAGATCGCC	1740
Db	1681	TGCGAAGCCAGAGCGGATCGCGGCTGTGACCCCTGTCCAAAGGCCAGTGTGGGCCAGATCGCC	1740
Qy	1741	CTGAGCTACGAGAAAGCTT 1758	
Db	1741	CTGAGCTACGAGAAAGCTT 1758	

```

RESULT 2
US-09-807-788-3
; Sequence 3, Application US/09807788
; GENERAL INFORMATION:
; APPLICANT: Stephen John Clifford Taylor
; APPLICANT: Robert Christopher Brown
; TITLE OF INVENTION: AMINOACYLASE AND ITS USE IN THE PRODUCTION OF D-AMINOACIDS
; FILE REFERENCE: GDE-239
; CURRENT APPLICATION NUMBER: US/09/807,788
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/GB99/03458
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1473
; TYPE: DNA
; ORGANISM: Alcaligenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1458)
US-09-807-788-3

Query Match
Best Local Similarity 81.9%; Score 1440.6; DB 31; Length 1473;
Matches 1449; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 31 TCCATGTCCTCATTCGATTCCTCCAGCCCTTGACGCTGTCTCGCGGGGCGACCCCTCATC 90
D 4 TCCATGCGCGGATCCGATTCCTCCAGCCCTTGACGCTGTCTCGTGTGGGGCGACCCCTCATC 63
QY 91 GAGGGCAGCAACACCCCGGGGGGGCGCGCCGACCTGTGGCGGTTCGCGGGGACCGCATCGCC 150
D 64 GAGGGCAGCAACACCCCGGGGGGGCGCGCCGACCTGTGGCGGTTCGCGGGGACCGCATCGCC 123
QY 151 GCGCATCGGCGATCTGTGGGAGCGCGCGCGCGGCAACCCGGGTGTGAGTGTTCGGGCTGTGGT 210
D 124 GCGATCGGCGATCTGTGGGAGCGCGCGCGCGGCAACCCGGGTGTGAGTGTTCGGGCTGTGGT 183

```

Qy	211	GGGGGGGGGGCTTAACTCAGACTGGCACACCCAGACGACAACTACCTGCTCAGGGCGTCG	270
Db	184	GTGGGGGGGGGGCTTCAATCAGCTCGCACACCCAGACGACAACTACCTGCTCAGGGCGTCG	243
Qy	271	GACATGACGGCCAAAGATCTTCGCGAGGGCGTCACACGAGGGTGTACGGGGCAATTGGGGATC	330
Db	244	GACATGACGGCCAAAGATCTTCGCGAGGGCGTCACACGAGGGTGTACGGGGCAATTGGGGATC	303
Qy	331	AGCGTGGCGGGCTGGCGCGACGCGAACCCGGCGGGCTTGGACCTGTGTGACGAAAGC	390
Db	304	AGCGTGGCGGGCTGGCGCGACGCGAACCCGGCGGGCTTGGACCTGTGTGACGAAAGC	363
Qy	391	GGCTCTTACCGTTTCAGAGCGTTTCGGCGACTACCTGAGAGGGTTGGGGGACAGCGGGG	450
Db	364	GGTTCGTAACCGGTTTGAAGGGCTTCCCGCACTCTGTGAGAGCGTTTGGGGGACAGCGGGG	423
Qy	451	GGCGTCAAGCGCGCGCTGTATGTGTGGCCATTCAACGCTGCGCGCGCGGGTATGTCCGGAC	510
Db	424	GGCGTCAAGCGCGCGCTGTATGTGTGGCCATTCAACGCTGCGCGCGGGTATGTCCGGAC	483
Qy	511	TTTCAGACGGCGCGCCACACGACGAGAAATTCGGGGCATATGGGGGACMTGGGCGAGGAAGC	570
Db	484	TTTCAGACGGCGCGCGCCACACGACGAGAAATTCGGGGCATATGGGGGACMTGGGCGAGGAAGC	543
Qy	571	ATGGCCAGGGGGCCCATGTGGCATTTTCAGCGGGCGCTTTCACCGCGCGCGCGCGCGCG	630
Db	544	ATGGCCAGGGGGCCCATGTGGCATTTTCAGCGGGCGCTTTCACCGCGCGCGCGCGCGCG	603
Qy	631	ACGACGAGAGATATCATGAGGTGTGGGGCGCGGTGAGAGCGCGCATGGCGGCATCTAGCC	690
Db	604	ACGACGAGAGATATCATGAGGTGTGGGGCGCGGTGAGAGCGCGCATGGCGGCATCTAGCC	663
Qy	691	ACCCACATCTCGCGACGAAAGCGGACACATCTGTGGCGCGCTGTGAGGAAACTTTCGCACT	750
Db	664	ACCCACATCTCGCGACGAAAGCGGAGACATCTGTGGCGCGCTGTGAGGAAACTTTCGCACT	723
Qy	751	GGCGCGAGACTGGAGTGCAGGTGTGGTGTATCTCGACCAAGATCATGGGCGACGCCAAT	810
Db	724	GGCGCGAGACTGGAGTGCAGGTGTGGTGTATCTCGACCAAGATCATGGGCGACGCCAAT	783
Qy	811	TTTCGGCGCTCGCGCGAGACGCTCCCGCTGATCTGAGGCGCGCATGTGGCGCGACGACGTC	870
Db	784	TTTCGGCGCTCGCGCGAGACGCTCCCGCTGATCTGAGGCGCGCATGTGGCGCGACGACGTC	843
Qy	871	TGCGTGGAGGGTAACTCTAGTGTGGCGGGCTCCACCATGCTCAAGCAGAACCGCGTGTG	930
Db	844	TGCGTGGAGGGTAACTCTAGTGTGGCGGGCTCCACCATGCTCAAGCAGAACCGCGTGTG	903
Qy	931	CTGGCGGAGCGACCATCATCATCTGTGTGAAGCGCTTCCCGCAACTGAGGGGGGGCAC	990
Db	904	CTGGCGGAGCGACCATCATCATCTGTGTGAAGCGCTTCCCGCAACTGAGGGGGGGCAC	963
Qy	991	CTGGATGAAAGTTCGGCGCGAGCGGGGCAATCCAAAGTACGAGGTGTGCCGAGCTGCAG	1050
Db	964	CTGGATGAAAGTTCGGCGCGAGCGGGGCAATCCAAAGTACGAGGTGTGCCGAGCTGCAG	1023
Qy	1051	CGGGCGGGCGCATCTACTCATGATGAGGAGAACCGGAGGTCAACGCAATCTGTGGCTTC	1110
Db	1024	CGGGCGGGCGCATCTACTCATGATGAGGAGAACCGGAGGTCAACGCAATCTGTGGCTTC	1083
Qy	1111	GGCCGACCATGATCGGCTCCGAGAGGCGTCCGCAAGACGACGCGCCGATCTCGGGCGCT	1170
Db	1084	GGCCGACCATGATCGGCTCCGAGAGGCGTCCGCAAGACGACGCGCCGATCTCGGGCGCT	1143
Qy	1171	TGGGGCACTTCCCGCGGGGTGTGGGGCACTATGGCGCGCACTGGGCGCTTCCCGCTG	1230
Db	1144	TGGGGCACTTCCCGCGGGGTGTGGGGCACTATGGCGCGCACTGGGCGCTTCCCGCTG	1203
Qy	1231	GAGAGCGCGATGAGAGATGACCGGCGTGAACCGCGCGCGGCTTGGCGTGGCGGGCGCG	1290
Db	1204	GAGAGCGCGATGAGAGATGACCGGCGTGAACCGCGCGCGGCTTGGCGTGGCGGGCGCG	1263

QY 1291 GGGCAGCTGACAGCGCGGGTACTTCCGACCTGGTGTGACACCGCCGACCGTGGCC 1350
 DB 1264 GGGCAGCTGACAGCGCGGGTACTTCCGACCTGGTGTGACACCGCCGACCGTGGCC 1323
 QY 1351 GATACCGGCACTTGGAAACACCTTACGAGCGCGCGCGGACATCATTCCTGTAGCTC 1410
 DB 1324 GATACCGGCACTTGGAAACACCTTACGAGCGCGCGGACATCATTCCTGTAGCTC 1383
 QY 1411 AACGGCGCGCGGTCTGTGCAAGACAGCGGCTTACCGGACGATGCGCGCGGTGCTC 1470
 DB 1384 AACGGCGCGCGGTCTGTGCAAGACAGCGGCTTACCGGACGATGCGCGCGGTGCTC 1443
 QY 1471 GCAGCGACGCGCGCTTACGCGCG 1493
 DB 1444 GCAGCGACGCGCGCTTACGCTCG 1466

RESULT 3
 US-09-489-039A-1847
 / Sequence 1847, Application US/09489039A
 / GENERAL INFORMATION:
 / APPLICANT: Gary Breton et. al
 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 / FILE REFERENCE: 2709, 2004.001
 / CURRENT APPLICATION NUMBER: US/09/489, 039A
 / PRIOR FILING DATE: 2000-01-27
 / PRIOR APPLICATION NUMBER: US 60/117,747
 / PRIOR FILING DATE: 1999-01-29
 / NUMBER OF SEQ ID NOS: 14342
 / SEQ ID NO 1847
 / LENGTH: 1479
 / TYPE: DNA
 / ORGANISM: Klebsiella pneumoniae
 / US-09-489-039A-1847

Query Match 24.4%; Score 428.6; DB 18; Length 1479;
 Best Local Similarity 57.5%; Pred. No. 3,4e-56;
 Matches 789; Conservative 0; Mismatches 579; Indels 3; Gaps 1;

QY 59 TCGACCTGCTGCTGCGCGCGGACCGCTCATCGACGACCAACCGCGCGCGCG 118
 DB 47 TTGACTGCTTTTCAAAAACGTCAGGTGATGACGAGGCGCGCGCGCAATACCGTG 106
 QY 119 CCGACCTGCGGCGTGGCGGCGGACCGATCGCGCGATCTGTGAGACGCGCGG 178
 DB 107 CTGATGCGGCGGTGAGAGGCGGATGATATGCTATCGCCCGCGCGGTGAGACGTTGCGG 166
 QY 179 CGACACCGCGGTGACGCTGTGCGGCGCTGTGCGCGCGCGCTTCACTGACGACCA 238
 DB 167 CGAGACGAGTGTATGACGAGGCGCGGCTACTGCGCGCGGCTTATGATGACATA 226
 QY 239 CCGACGACGACAACTACCTGCTGACGCGCTGCGGACATGACGCGCAAGATCTGCGAGGCG 298
 DB 227 CCGATGATATATCAACGCGATCGCATGCGGAGTACTTGGCAAGCTCACCGACGAGGCG 286
 QY 299 TCACACGAGTGTGACGCGCAATTCGGCATACGCTGCGCGCGCGTGC---GCACGCGCA 355
 DB 287 TGACACGAGTGTGCTGCGCAACTGCGGATGATGCGCGCGCGCGCACACATGCGGCGG 346
 QY 356 ACCGCGCGCGCGCTGACCTGCTGACGAGAGCGGCTTACCGTTGAGCGCTTGC 415
 DB 347 AAGTGGCGGACCGCATGATCTCTCGGCGAGCAACACTTATATCCACCGCTG 406
 QY 416 CCGACTACCTGACGCGGTGCGGCGGACCGCGCGCGCTTCAACGCGCGCTGTATGCTG 475
 DB 407 AGGCTTATGCGGACCGGTGAGGCGGCGGCGCGCGCTGCGTGAATGCGGACCTTATG 466
 QY 476 GCCATTACGCTGCGCGCGGTGCTGCGGACTTGGACGCGCGCGCGCGCGAGG 535
 DB 467 GTCAACCGCGCGGTGCTGATATACATATGACATCTGTTCTGCGGCGGAGACGAG 526
 QY 536 AATTCGCGCGCGGACCTGCGCGGAGACGAGCAATGCGAGCGCGCGCATCGGATTT 595

DB 527 AATTCGCGCGCGGATGCGGATACACTACGCGATCGCTGCGCGAGGAGATTAAGGCTCA 586
 QY 596 CGACCGCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 655
 DB 587 GTACCGCGCGGTGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 646
 QY 656 GCGCGCGCGGTGAGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 715
 DB 647 CGAAGAGTGGCG 706
 QY 716 ACATGCTGCG 775
 DB 707 CGATTCTGAGCG 766
 QY 776 TGATCTGACACGACGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 835
 DB 767 TGCTTCCGACGACGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 826
 QY 836 CGCTGATGAGCG 895
 DB 827 CTTTTCGATGAGATGCG 886
 QY 896 CCGGCTCCACCATGCTCAAGCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 955
 DB 887 CCACTGCTGACGCGCTGATATGACGAGGTGACCGCGCGCGCGCGCGCGCGCGCGCG 946
 QY 956 GGTGCAAGCGCTTCCCGGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
 DB 947 GTTCGAGGCG 1006
 QY 1016 GCAATTCAGATACGACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075
 DB 1007 AGGTGACGCTGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1066
 QY 1076 TGACAGACCG 1135
 DB 1067 TGATGAGAGAGACGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1126
 QY 1136 GCGTCCGCGACGACG 1195
 DB 1127 GCGTCCGCGACGATCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1186
 QY 1186 GGCATATGCG 1255
 DB 1187 GCGCATATGCG 1246
 QY 1256 GCGTACCG 1315
 DB 1247 GCGTGTGCG 1306
 QY 1316 CCGACCTGCGGCGGTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1375
 DB 1307 CCGACCTGCGGCGGTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1366
 QY 1376 CCGAGCG 1426
 DB 1367 AACGACG 1417

RESULT 4
 US-60-184-777-35
 / Sequence 35, Application US/60184777
 / GENERAL INFORMATION:
 / APPLICANT: Hodgson, David M.
 / APPLICANT: Lincoln, Stephen E.
 / APPLICANT: Russo, Frank D.
 / APPLICANT: Spillo, Peter A.
 / APPLICANT: Banville, Steve C.
 / APPLICANT: Bratcher, Shawn R.
 / APPLICANT: Dufour, Gerard E.
 / APPLICANT: Cohen, Howard J.
 / APPLICANT: Rosen, Bruce
 Jones, Anissa L.
 Yu, Jimmy Y.
 Greenawalt, Lila B.
 Panzer, Scott R.
 Roseberry, Ann M.
 Wright, Rachel J.
 Chen, Mensheng
 Liu, Tommy
 Yap, Pierre E.

[illegible][illegible]


```

OTHER INFORMATION: ORF 29; positive strandedness
NAME/KEY: misc_feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
NAME/KEY: misc_feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
NAME/KEY: misc_feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; positive strandedness
NAME/KEY: misc_feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-976-059-1

```

```

Query Match      4.9% Score 86; DB 36; Length 88421;
Best Local Similarity 43.1%; Pred. No. 0.00037;
Matches 603; Conservative 0; Mismatches 775; Indels 24; Gaps 3;

```

```

QY 35 TGTCCCAATCCGATTCACGACCTTTCGACCTGCTGCGGCGGCGACCCCTCATCGACG 94
DB 46230 TGGCCCGGACAGAGTGTGTGCAAGGTGTTCAGGTGCGGACGCGACCGCTGCGCAC 46289
QY 95 GCACACACACCCGCGGCGGCGCGCCGACCTGCGGCGTGGCGGCGACCGCATCGCCGCA 154
DB 46290 ACATCTCTGACCCCGCGGCGCGGCGCGGCTGCGGCTGTGAGAGTGGCGGCGGCG 46349
QY 155 TCGGCGATGTGTGAGAGCGCGCGCGCGACACCCCGGTGACGTGTGCGGCTGTGCTG 214
DB 46350 GGGTCGAGAGGCGGCGTGGCGGCGGCGGCGGCGGCTGCGGCTGTGAGAGTGGCG 46409
QY 215 CGCCGCGCTTCATGACCTGCGACACCGACGACGACACTACTGCTGAG-----GCCTC 268
DB 46410 CGGTGCGTGGCGGTGTCTGTGCGCCCGCGGCGACGCGACCGCATCTGCTGTGCTGCG 46469
QY 269 GCGACATGACCGCCAGATGTCGAGGCGGCTGACACAGGTGTGACGCGGCAATTGCGCA 328
DB 46470 ACCACATGCGGCGCGAGGCGGTGTGATGCGGCGGCTGCGGCGGCGGAGTGGACCTG 46529
QY 329 TCAGCCGTGGCGCGGCGTGGCGACCGCAACCGCGCGCGCGCGCTGAGCTGTGACGAG 388
DB 46530 ACGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46589
QY 389 GCGGCTCTTACCGTTTCGACGCGCTTCGACCTGACGCGCTTTCGCGGCGACGCGCG 448
DB 46590 ACTACGCGCTCTGACGCGTGTGCTGTGCGGCTGCGACGCGACCGCGGCGGCGGCTG 46649
QY 449 CGGCGCTGACGCGCGCTGTATGTGCGGCGGCTGACGCGTGTGCGGCGGCGGCTGCGG 508
DB 46650 CCGGCGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46709
QY 509 ACTGAGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 568
DB 46710 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46769
QY 569 CCATGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 628
DB 46770 TCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 46829
QY 629 CCACACCGGAGGATGATGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 688
DB 46830 TCATGACCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46889
QY 689 CCACCCATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 748
DB 46890 TCCGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46949
QY 749 TGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 808
DB 46950 TCTGTGTAACGCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 47009
QY 809 ATTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 868

```

```

DB 47010 TGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47069
QY 869 TCTGCGTGGAGCGGCTATTCCTTACGTGCGCGGCTCCACCATGCTCAAGAGAGCGCGTGC 928
DB 47070 AGAAGCTGTGTGAGAGAGTCAACCCCGGCGGCTC-----GCTGCGCGGCGGCGGCTGT 47123
QY 929 TGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 988
DB 47124 TCCAGGTATGTGTCACTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47183
QY 989 ACCGTGATGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1048
DB 47184 CGATGTGCTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47243
QY 1049 AGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1108
DB 47244 CCGAGACCTTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47303
QY 1109 TCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1168
DB 47304 CCGACCTTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47363
QY 1169 TGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1228
DB 47364 GCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47411
QY 1229 TGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1288
DB 47412 TGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 47471
QY 1289 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1348
DB 47472 CCGGAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47531
QY 1349 CCGATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1408
DB 47532 CGGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47591
QY 1409 TCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1432
DB 47592 TCGCGCGCTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47615

```

```

RESULT 7
US-09-739-449-925
Sequence 925, Application US/09739449
GENERAL INFORMATION:
APPLICANT: Hinkle, Gregory J.
TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15490)C
CURRENT APPLICATION NUMBER: US/09/739,449
PRIOR FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: US 09/514,000
NUMBER OF SEQ ID NOS: 13351
SEQ ID NO 925
LENGTH: 770
TYPE: DNA
ORGANISM: Agrobacterium tumefaciens
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(770)
OTHER INFORMATION: unsure at all n locations
US-09-739-449-925

```

```

Query Match      4.9% Score 85.4; DB 29; Length 770;
Best Local Similarity 51.0%; Pred. No. 0.00089;
Matches 214; Conservative 0; Mismatches 205; Indels 1; Gaps 1;
QY 1078 GACGACCGGAGGTGAGAGGCGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1137
DB 35 GANCAACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94

```

QY 1138 CTGCGGACAGAGAGAGCCGCGGATCCGCGCTGTGGGACACTTCCGCGGGTGGGG 1197
 Db 95 GGG 154
 QY 1198 CACTGTGGGCGGAGACTGTGGCTGTCCCGTGGAGAGCGGGATATGAGATGACCGGC 1257
 Db 155 CCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 214
 QY 1258 CTGACCGCGCGGCTTGGCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1317
 Db 215 CCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 274
 QY 1318 GACTGTGGTGTTCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1377
 Db 275 CCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 334
 QY 1378 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1437
 Db 335 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 394
 QY 1438 GCGTTACCGCGGACGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1497
 Db 395 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 453

RESULT 8

US-09-803-110-925
 ; Sequence 925, Application US/09803110
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15490)D
 ; CURRENT APPLICATION NUMBER: US/09/803,110
 ; PRIOR FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 09/739,449
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: US 09/514,000
 ; PRIOR FILING DATE: 2000-02-23
 ; PRIOR APPLICATION NUMBER: US 60/168,139
 ; PRIOR FILING DATE: 1999-12-01
 ; NUMBER OF SEQ ID NOS: 13351
 ; SEQ ID NO 925
 ; LENGTH: 770
 ; TYPE: DNA
 ; ORGANISM: Agrobacterium tumefaciens
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(770)
 ; OTHER INFORMATION: unsure at all n locations
 US-09-803-110-925

Query Match

Best Local Similarity 4.9%; Score 85.4; DB 31; Length 770;
 Matches 214; Conservative 0; Mismatches 205; Indels 1; Gaps 1;

QY 1078 GACGAACCGGACGCTGACGAGCTGAGGCTTGGGCGGCGGCGGCGGCGGCGGCGG 1137
 Db 35 GAANAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94
 QY 1138 CTGCGGACAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1197
 Db 95 CCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 154
 QY 1198 CACTGTGGGCGGAGACTGTGGCTGTCCCGTGGAGAGCGGGATATGAGATGACCGGC 1257
 Db 155 CCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 214
 QY 1258 CTGACCGCGGCGGCTTGGCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1317
 Db 215 CCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 274

QY 1318 GACTGTGGTGTTCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1377
 Db 275 CCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 334
 QY 1378 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1437
 Db 335 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 394
 QY 1438 GCGTTACCGCGGACGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1497
 Db 395 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 453

RESULT 9

US-09-758-759-1
 ; Sequence 1, Application US/09758759
 ; GENERAL INFORMATION:
 ; APPLICANT: Hosted, Thomas J.
 ; APPLICANT: Hangan, Tim X.
 ; APPLICANT: Horan, Ann C.
 ; TITLE OF INVENTION: EvernInomleIn Biosynthetic Genes
 ; FILE REFERENCE: ID0983K US
 ; CURRENT APPLICATION NUMBER: US/09/758,759
 ; CURRENT FILING DATE: 2001-01-11
 ; PRIOR APPLICATION NUMBER: US 60/175,751
 ; PRIOR FILING DATE: 2000-01-12
 ; NUMBER OF SEQ ID NOS: 204
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 109519
 ; TYPE: DNA
 ; ORGANISM: Micromonospora carbonacea
 US-09-758-759-1

Query Match

Best Local Similarity 4.8%; Score 84.6; DB 29; Length 109519;
 Matches 542; Conservative 0; Mismatches 629; Indels 30; Gaps 5;

QY 117 GCGGACCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 176
 Db 94498 CGTGACGCTGTTCACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94557
 QY 177 GCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 236
 Db 94558 GAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94617
 QY 237 CACCCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 296
 Db 94618 CGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94677
 QY 297 CGTACACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 356
 Db 94678 GGTACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94737
 QY 357 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 404
 Db 94738 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 464
 QY 405 CGAGCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 524
 Db 94738 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 584
 QY 465 CTGTATGTGGCCATTACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94917
 Db 94858 GCGCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94977
 QY 525 CACGAGAGAGAAATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 584
 Db 94918 CTTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94977
 QY 585 CATCGGCAATTTCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 644
 Db 94978 CCGGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 95037

QY 645 CATGAGGTGTGCGGCGCGGTGAGCCGCGCATGCGGCATTTACGCCACCCACATGCGCGA 704
 Db 95038 CACCTCCGCGCGCGGTCTCTACCGCGCACGGTGCACCCGAGCGCGGCTCACCTCGA 95097
 QY 705 CGAAGCGAGACATGCTGCGCGCGGTGAGAGAACTTCCGATGGCGCGCGAGTGA 764
 Db 95098 CGGAGCGAGCTGTGCGCGGTGAGAGAACTTCCGATGGCGCGCGAGTGA 764
 QY 765 CGGCGCGGTGTGATCTGCGAGCACAAGTTCATGCGCGCGCGCGCGCGCGCGCGCG 818
 Db 95158 CGCGAGCGGTGTGATCTGCGAGCACAAGTTCATGCGCGCGCGCGCGCGCGCGCG 818
 QY 819 CTGCGCGCGAGCTGTGCGCGGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95217
 Db 95218 CCGCGCGCGAGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95217
 QY 879 GCGGTATCCCTGATGCGCGCGGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 934
 Db 95278 CCGCGCGCGAGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95337
 QY 935 CCGAGCGCGACATCATCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 992
 Db 95338 CCGAGCGCGAGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95397
 QY 993 GATGATGCTGCG 1046
 Db 95398 CGAGCGCGCGAGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95457
 QY 1047 GCAGCG 1106
 Db 95458 GCG 95517
 QY 1107 GTTGGCG 1166
 Db 95518 CGACCTGCGAGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1166
 QY 1167 CCGTGGCG 95577
 Db 95578 CGATTGCG 1226
 QY 1227 GCTGAGCGCGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95637
 Db 95638 GCGCGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1286
 QY 1287 G 1287
 Db 95698 G 95698

RESULT 10
 US-09-758-759-150
 ; Sequence 150, Application US/09758759
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tim X.
 ; APPLICANT: Hosted, Thomas J.
 ; APPLICANT: Horan, Ann C.
 ; TITLE OF INVENTION: Evernolmclin Biosynthetic Genes
 ; FILE REFERENCE: ID0983K US
 ; CURRENT APPLICATION NUMBER: US/09/758,759
 ; CURRENT FILING DATE: 2001-01-11
 ; PRIOR APPLICATION NUMBER: US 60/175,751
 ; PRIOR FILING DATE: 2000-01-12
 ; NUMBER OF SEQ ID NOS: 204
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 150
 ; LENGTH: 1839
 ; TYPE: DNA
 ; ORGANISM: Micromonospora carbonacea
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1839)
 ; OTHER INFORMATION: evce

US-09-758-759-150
 Query Match 4.8%; Score 84; DB 29; Length 1839;
 Best Local Similarity 45.6%; Pred. No. 0.0013;
 Matches 515; Conservative 0; Mismatches 555; Indels 30; Gaps 5;
 QY 117 GCGCGACACCGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 176
 Db 669 CGTGAACCTTGTTCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728
 QY 177 CCGCGACACCGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 236
 Db 729 GAGACCGCGCGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788
 QY 237 CAGCGAGACAGCACTTCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296
 Db 789 GCGCTTCCCG 848
 QY 297 GGTACACAGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
 Db 849 GGTACACAGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 908
 QY 357 CCG 404
 Db 909 CCG 968
 QY 405 CGACCGGTGTGAGAGCG 464
 Db 969 CCG 1028
 QY 465 CTGTATGCG 524
 Db 1029 CCG 1088
 QY 525 CAGCGAG 584
 Db 1089 CTTGAGAGCGGTGTGAG 1148
 QY 585 CATGCG 644
 Db 1149 CTTGAGAGCGGTGTGAG 1208
 QY 645 CATGAGGTGTGCG 704
 Db 1209 CAGCG 1268
 QY 705 CGAAGCGAGACATGATGCG 764
 Db 1269 CCG 1328
 QY 765 CCGCGCGGTGTGATCTGCGAGCACAAGTTCATGCGCGCGCGCGCGCGCGCGCGCGCG 818
 Db 1329 CCGCGCGGTGTGATCTGCGAGCACAAGTTCATGCGCGCGCGCGCGCGCGCGCGCGCG 1388
 QY 819 CTGCGCGAGAGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 878
 Db 1389 CCGCGAGAGAGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1448
 QY 879 CCGGTATCCCTGATGCGCGCGGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 934
 Db 1449 CCGCGAGCGCGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1508
 QY 935 CCGAGCGACATCATCATGCGCGCGGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 992
 Db 1509 CCGAGCGAGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1568
 QY 993 GATGATGCTGCG 1046
 Db 1569 CGAGAGCGCGAGCGGTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1628
 QY 1047 GCAGCG 1106
 Db 1629 GCGCGAGACCG 1688

QY 1107 GTTGGCCGACCATGATGCGCTCCGACGCGCTGCGGACGAGGCGCCGCAATCCGCG 1166
 Db 1689 CGACCTCGACGAGATGCGTGTCTCGACGCGCGGCGGTGTCTCCAGCGGTGCGGACAGA 1748
 QY 1167 CCGTGGGCGACCTTCCCGCGGCGTGTGGGGCACTATGCGCGGACCTGG 1216
 Db 1749 CGAGTGTGTGCGCGCGCGGCGTGTACCGGAGACATGCGTGTCTCCAGG 1798

RESULT 11 PCT-US02-08937-1/C

Sequence 1 Application PC/TUS0208937
 GENERAL INFORMATION:
 APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
 TITLE OF INVENTION: LEINAMICIN BIOSYNTHESIS GENE CLUSTER AND ITS COMPONENTS AND THEIR
 FILE REFERENCE: 309T-000110PC
 CURRENT APPLICATION NUMBER: PCT/US02/08937
 PRIOR FILING DATE: 2001-03-22
 PRIOR FILING DATE: 2001-03-26
 NUMBER OF SEQ ID NOS: 222
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 1
 LENGTH: 135638
 TYPE: DNA
 ORGANISM: Streptomyces atroolivaceus
 PCT-US02-08937-1

Query Match 4.7% Score 82; DB 1; Length 135638;
 Best Local Similarity 44.3%; Pred. No. 0.0014;
 Matches 610; Conservative 0; Mismatches 750; Indels 18; Gaps 6;

QY 117 CGCGACCTGGGCGCTGCGCGGACGCGCATGCGCGCATCGCGCATCTGTGGAGCGCGC 176
 Db 118012 CGACGAGCTCGCTCCCGACGCGCGCATCTCACCGGACCGGACCGCATCTGTGACGAG 117953
 QY 177 CGCGACACCGCGGCTCGACGCTGTGGCGCTGTGGCGCGCGCGCGCTTCATGACCTGCA 236
 Db 117952 CTGCGCGACCTGTGTGACGCGGAGTGGAGCGCGCGCGCGCATGCGCGCGCGCG 117893
 QY 237 CACCCAGACGAACTACCTGCTGACGAGCGCTGCGGACATGACGCGGACGCGGACGCG 296
 Db 117892 GCTGCGCGCGGACGAGCGCGGACCGGACCGGATGCGGATCTTCTTGGGACGAGCTG 117833
 QY 297 GCTACACGAGTGTGACGCGGCAATGCGGATGACGCGGCGCGCGCGCGCGCGCGCA 356
 Db 117832 GCGCGGCTACGCGGTGCTACCTCGCGCGCGCTCAAGCGCGCGCGCGCGCGCACT 117773
 QY 357 CCGCGCGCGCGCGCTGACCTGCTGAGAGAGCGCGCTTACCGTTTGAAGCGCTTCCG 416
 Db 117772 GCGCGTGGCGCTCGCGCGGACGAGCTGAGAGCGCGGTGCGCGCTGATGCGAGC---TGGC 117716
 QY 417 CGACTACCTGACGCGTTCGCGGCGGCGCGCGCGCGCTGACGCGCGCGCTGTATGGTGG 476
 Db 117715 CGGCGATGTCACGCGGCGGACCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGGAC 117656
 QY 477 CCAATTACGCTGCGCGCGCGGTATGCGGACTTGCAGCGCGCGCGCGCGCGCGCGCG 536
 Db 117655 GCTGCAAGAGCTGCTCGCGCGCGCGGAGACACCGGCTGACGCTGTCACAGCGCGCG 117596
 QY 537 AATCGCGCGCATGCGGAGCGCGCGGAGAGACGATGCGGACGCGCGCGCGCGCGCGCTTTC 596
 Db 117595 GCGCGCGCGGATGCTGCTCTCTCGGACACCGCGAGCTGCGCGCGCGCGCGCGCTGCTC 117536
 QY 597 GACCGCGCGCTTACCG 656
 Db 117535 GCACGAGAACTCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 117476
 QY 657 CC---GCGCGCGAGCG 713
 Db 117475 CCGCATGCTGTGCG 117416

QY 714 GCACATGTGTGCGCGCGCTGAGAGAACTTCCGATCGCGCGCGAGCTGAGACGCGGT 773
 Db 117415 GCTCAAGCG 117356
 QY 774 GGTATCTTCACACGAGGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 833
 Db 117355 CGGCGTATGAGACG 117296
 QY 834 GCGCGTATGAGCG 893
 Db 117295 GATGACCG 117239
 QY 894 GCG 953
 Db 117238 CACCG 117179
 QY 954 CTGTGCAAGCG 1013
 Db 117178 GCG 117119
 QY 1014 CGGCAATCAAGTACGAGCG 1073
 Db 117118 CTTCGACCG 117059
 QY 1074 GATGACGAGCG 1130
 Db 117058 CACGACGAGCG 116999
 QY 1131 CGAGCG 1187
 Db 116998 CGAGCG 116939
 QY 1188 GGTGCTGCG 1247
 Db 116938 GCGCGAGTGTGCG 116879
 QY 1248 GATGACG 1307
 Db 116878 CTTCGACG 116822
 QY 1308 GTACTTCCG 1367
 Db 116821 GGTGCGAGTGTGCG 116762
 QY 1368 ACACCGTACGAGCG 1427
 Db 116761 GCG 116702
 QY 1428 GCAGAGCAGCG 1485
 Db 116701 GCTGCG 116644

RESULT 12

US-09-836-821-7
 Sequence 7, Application US/09836821
 GENERAL INFORMATION:
 APPLICANT: Sherman, D.H.
 APPLICANT: Liu, H.
 APPLICANT: Xue, Y.
 APPLICANT: Zhao, L.
 TITLE OF INVENTION: DNA encoding methymycin and plitromycin
 FILE REFERENCE: 600.438US1
 CURRENT APPLICATION NUMBER: US/09/836,821
 PRIOR FILING DATE: 2001-04-17
 PRIOR FILING DATE: 1998-06-26
 NUMBER OF SEQ ID NOS: 43
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 7
 LENGTH: 1248
 TYPE: DNA
 ORGANISM: Streptomyces venezuelae

US-09-836-821-7

Query Match	4.7%	Score 81.8;	DB 32;	Length 1248;
Best Local Similarity	44.28;	Pred. No. 0.003;		
Matches 439; Conservative	0;	Mismatches 542;	Indels 12;	Gaps 2

QY	32	CCATGTCCTCCATCCGATATCTCCACACCCCTTGAGACTGCTGTCCGGGGCGGCACCTCATG	91
Db	215	CCACTCTCAAGCGACCGCGGGGCTCCAGCTCTCTCCGACCGCGCGCCCTCACCGGCG	27
QY	92	ACGGCACCAACACCCGGGGGGCGCGCGCAGCTGGCGCTGGCGGCGACAGCATGCGCG	151
Db	275	AAGGATCATGTCGGTGATATGATCTTGGCGCCACCCCGACACGACTTGGCTGATGGCC	334
QY	152	CCATGCGGATCTGTGTGGAGAGCGCGCGGCGACACCGGGTGACGTGTGCGGCTGGTG	211
Db	335	TCACCCCGGCTTCTGGCGACATTCAGCCCGAACCGGCACTCGACCCGGACACAGGTG	394
QY	212	TCGGGCCCGGCTTCAATGACTGCGACACCGACAGACAACTACTGCTCAGCGGTGCG	271
Db	395	CCGCGCGGGTCACACCCCGCACCTGGCGCGCTGTGTGGGTGCACCTCTGGGGCGCCCT	451
QY	272	ACATGACGCCCAAGATCTGCGACGGGGGTACACAGGTGTATACGGGCAATTGCGGCTCA	331
Db	455	GCGCGCGCGACGAGGTGCGGAAAGTGTGCGACGACAGCGCTGCGGTATCTTCACG	514
QY	332	GCTTGGCGCGCTGCGGCGACGCCAACCGCGCGCCCGCTTGAGACTGTGTGACGAAAGCG	391
Db	515	CCGCGACAGCCCTCGGCTGTGCGGCTGACAGCGCGCGCGCGGACGCTGGGGAACCGG	574
QY	392	GCTTTCACGTTTCGAGCGCTTCCGCGGACTACGTGAGACGGTTCGGGGCCACGCGCGG	451
Db	575	AGGTTCTTAGCTTCCA---CGCCACCAAGGCGCTTCAGCGCTTGAAGGGCGCGCGTGC	631
QY	452	CGGTAAAGCCCGCTGTATGTGTGGCCATTAAAGCTGTGCGCGCGCGCTATCGCGACT	511
Db	632	TCACGAGACGCGCGACCTCGCGCGCGGATCGCGCGCTTCACAACTTGGCTTGACG	691
QY	512	TCGACGCGCGCCCGCACGACGAGAAATCGCGGCGATGCGGGACCTTGGCGGAAAGAACCA	571
Db	692	TCGCGGCGCGCAGCGCCCGGGGGAACACAGCGCAAGTATGACGAGAGCGCGCGGCA	751
QY	572	TGGCCAGGGGGCCATTCGGGATTTCGACCGGCGCTTCTACCGCGCGCGCGCGCGCA	631
Db	752	TGGGCTTACCTCCCTCGACCGCTTCCCGAGGTCAATCGAACCGGAACCGCGCAACAG	811
QY	632	CCACGGAAGATCATCGAGGTGTGCGCGCGCTGAGCGGCGCATGGCGGATTTACGCCA	691
Db	812	CGGCTTACCGGAGACCTCGCGGACCTCCCGCGCTCTCTGTGCGCGACACAGACCGCG	871
QY	692	CCCAATCGCGAGAGGCGACGACATCTGTGCGCGCGCTGAGGAACCTTCCGCACT	751
Db	872	ACGGCTTAAACACACAGTACGTGATCGTCGATCGATCGACGAGGACACACACCGGATGC	931
QY	752	GCGGAGACTGAGAGTGCGGGTGTGATTCGGAACCAAGATCATGGGCGACGCCAATT	811
Db	932	ACCGGAGACTGTATGAGAGTCTCTGAAGCGGAAAGCGTGCACACCGCGGCTACTTCT	991
QY	812	TCGGCGCTCGCGGAGACGCTGCGGCTGATCGAG-----GCCGCAATGGCGCGCG	862
Db	992	CGCGGGCTGCGACGACTGAGACCGGACCGGACGCGGACGCGACCGCCCGCTGCGGCA	1051
QY	863	AGGAGCTCTGCGTGGAGCGGCTATTCCTAGTGGCGCGGCTTCACATGCTCAAGCAGACC	922
Db	1052	CCGAAGCGTCTGCGCGCGCGCGGTCTCTCCCTGCGGACCGGACCGCAATCGCGCAGCG	1111
QY	923	GCGTGTCTGTGGCGCGACGCAACATCATCACTGTGTGAAGGCTTCCCGGAATGAGCG	982
Db	1112	ACATTCGCCCGGTGCGGACCTGTCTGTCTGTGCGCGAACCGCGGCGCGCAACTGACCG	1171
QY	983	GCGCGACCTGATGAAGTGGGGCGGAGCGGG	1015
Db	1172	CGGCGCACCGGACGCGCGCGCGCGCGCTGG	1204

RESULT 13
US-09-860-846-7

```

sequence /, Application US/09860846
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: xue, y.
APPLICANT: xue, y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and plitromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/860,846
PRIORITY FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIORITY FILING DATE: 1998-05-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 1248
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-860-846-7

```

Query Match	4.7%	Score 81.8;	DB 33;	Length 1248;
Best Local Similarity	44.2%;	Pred. No. 0.003;		
Matches 439;	Conservative 0;	Mismatches 542;	Indels 12;	Gaps 2

QY	32	CGATGTCGCAATCGGATTCGAGCCCTTCACCTGCTGCAGTGGCGGGCGGACACCTTCATCG	91
Db	215	CGACCTTCAACAGCGCACGCGCGGGCTTCAGCTCTCTCGGGACAGCGCGCGCGCTTCACGGCG	274
QY	92	ACGGCAGCAACACACCGCGGGCGGGCGGCGCAACTGGGCGGTGCGCGGACACCGCATCGCG	151
Db	275	AAGGATCATGACCGCTGATGATGATGCTGGCGGCCACCCCGCAGCATGCGCTGATGCGGC	334
QY	152	CGATCGGCGATCTGTGTGGAGCGCGCGCGCACCCCGGGTGCAGTGTGGCGCTGTGGT	211
Db	335	TCACCCCGGGCTTCGCGCGCATGACCGCGGACACCGGCAACCTGCAGCCGGACAGATGG	394
QY	212	TCACGCGCGGCTTCATCGATCGTGCACACCGACAGAGACAACTACCTGCTCAGAGCTGGCG	271
Db	395	CGCGCGGGGTACACACCCCGCACCTGCGGGCTGTGTGGCTCTCACTGTGGGCGCCCTT	454
QY	272	ACATGACGCCCAAGATCTGTGCGAGGGCGTGCACACGATGTGTCACGGGCAATTGCGGCATCA	331
Db	455	GGCGCCCGCCACAGCTGCGGAGAGGTGCGCGAGAGAGAGCGCTGGGGCTGTCTTCGACG	514
QY	332	GCCTGCGCGCGGTGGGGCGACAGCAACCGCGCCCGCCCTTGGACTGCTGACGAGAGCG	391
Db	515	CGCGCGACGCCCTCGGCTGCGGCTGAGAGCGCGCGCGCGCGCGACCTTCGCGGACGCGG	574
QY	392	GCTCTTACCGTTTGAAGCGCTTCGCGCATCACTGAGACGCTTGGCGGGCGCAAGCGGGGG	451
Db	575	AGGTCTTCAGCTTCCA---CGCCACCAAGGGCGGTGCAGAGCGCTTCGAGGGCGGGCGCGTGG	631
QY	452	CGGTACCGCGCGCGTGTATGTGGGGCATTTACAACGCTGCGGCGCGGCTCATGCGGCACT	511
Db	632	TCACGAGAGACGCGCACTCGCGCGCGGATTCGCGCGCTTCACAACTTTCGGCTTTCGAC	691
QY	512	TGCAAGCGGCGCGCCACCGACGAGAAATCGGCGGCATCGGAGACTGGGCGGAGGAGCA	572
Db	692	TGCGCGGGGCGGACGCCCGCGCGGGGACCAAGCCAAATATAGGAGGCGCGCGCGCA	751
QY	572	TGGCGAGGGGGCGCATTCGGCATTTTCGACCGCGCGCTTTCATCCGCGCGCGCGCGCGCA	631
Db	752	TGGGCTTACTCTCCCTGCACGCTTTTCCCGAGGTCAATCGACCGGAAACCGGCGCAACAG	811
QY	632	CGACGAGAGATCATCGAGGTGTGCGCGCGCTGAGACGCGGCATGGGGGATGTCAAGCCA	691
Db	812	CGCGCTTACCGGACACACTCGCGGACCTCCCGCGCGTCTGTGTGCGGACACGACGCGCC	871
QY	692	CGCAAGTCCGGGAGAGAGGAGCACTACTGTGGCGCGGCTGGAGAAAACCTTCGCGATCG	751

Db 872 ACGGCTTCACACACACAGTACGTGATGCTGAGATGACGAGGACACACCGGATTC 931
QY 752 GCGGAGCTGAGAGCTGCGGCTGATCTGACACAGTATGAGGACAGCCATT 811
Db 932 ACCGACCTGCTATGAGAGTCTGAGAGGACGAGGACAGCCGCTTCT 991
QY 812 TCGGCGCTGCGGAGAGCTGCTGATGAG-----GCCGACATGCGGCGC 862
Db 992 GCGGCGCTGCGGAGAGCTGAGAGCTGAGAGGAGGAGGAGGAGGAGGAG 1051
QY 863 AGGAGCTGCTGAGAGGCTATGCTGAGAGGAGGAGGAGGAGGAGGAGGAG 922
Db 1052 CCGAGAGGCTGCGGCGGCGGCTGCTGCTGCGGAGAGGAGGAGGAGGAGGAG 1111
QY 923 GCGTGTGCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 982
Db 1112 ACATCGCGGCGGCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1171
QY 983 GCGGAGCTGATGAGTGAAGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1015
Db 1172 CCGGCGACCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1204

RESULT 14
US-09-861-289-7
; Sequence 7, Application US/09861289
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and plikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-7

Query Match
Best Local Similarity 44.7%; Score 81.8; DB 33; Length 1248;
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

QY 32 CCATGTCCTCATCCGATTCCTCAGGCTTCTGACCTGCTGCGGCGGAGCCTCATCG 91
Db 215 CCACCTTCACACGCGGCGGCTGCTGCTGCGGAGGCGGCGGCTCATCGGCG 274
QY 92 ACGGAGCAACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 151
Db 275 AAGTGTATCATGCGTGTGATGAGCTTGGCGGCGGCGGCGGCGGCGGCGGCGG 334
QY 152 CCATGCGGCTGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 211
Db 335 TCACCGCGGCTTCTGCGGAGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 394
QY 212 TCGCGCGGCTTCTGAGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 271
Db 395 CCGCGCGGCTTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 454
QY 272 ACATGAGCGGCGGAGATCTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 331
Db 455 GCGCGCGGCGGAGCTGCGGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 514
QY 332 GCGTGTGCGGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 391
Db 515 CCGCGGAGCGGCTGCGGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 574

QY 392 GCTCTTACCTTTGAGAGCGCTTCCGAGCTACCTGAGAGGCTTCCGCGGACCGCGCG 451
Db 575 AGGCTTACCTTCA-----GCCACCAAGGCGCTTCAAGGCTTGAAGGCGGCGGCG 631
QY 452 CCGTCAACCGCGCTGTATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 511
Db 632 TCACCGAGAGCGGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 691
QY 512 TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 571
Db 692 TCGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 751
QY 572 TCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 631
Db 752 TCGGCGCTACCTTCCGAGAGGCTTCCGAGAGTATGAGAGGAGGAGGAGGAGG 811
QY 632 CCACCGAGAGATATCGAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 691
Db 812 CCGCTTACCGGAGAGCTTCCGAGAGTCCCGGCGGCTGCTGCTGCGGCGGAGCGG 871
QY 692 CCACATGCGGCGGAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 751
Db 872 ACGGCTTCAACACACAGTACGATGATGAGAGGAGGAGGAGGAGGAGGAGGAG 931
QY 752 GCGGAGCTGAGCTGCGGCTGATCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 811
Db 932 ACGGAGCTGATGAGAGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 991
QY 812 TCGGCGGCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 862
Db 992 CCGGCGGCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1051
QY 863 AGGAGCTGCTGAGAGGCTATGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 922
Db 1052 CCGAGAGGCTGCGGCGGCGGCTGCTGCTGCGGAGAGGAGGAGGAGGAGGAG 1111
QY 923 GCGTGTGCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 982
Db 1112 ACATCGCGGCGGCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1171
QY 983 GCGGAGCTGATGAGTGAAGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1015
Db 1172 CCGGCGACCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1204

RESULT 15
US-09-988-384B-7
; Sequence 7, Application US/09988384B
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and plikromycin
; FILE REFERENCE: 600,536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-7

Query Match
Best Local Similarity 44.7%; Score 81.8; DB 37; Length 1248;
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

QY 32 CCATGTCCCATCCGATTCCTCCAGCCCTTGACCTGCTGCTGGCGGGGGACACCCCTCATCG 91
Db 215 CCACCTGCAGACGCAAGGGCGGGGCTCCAGCTCTCGCGCAAGCGCGCTTCACCGGGCG 274
QY 92 ACGGAGCAACACCCCGGGGGCGGGCGGACCTGGGCGTGGCGGGCGACCCGATCGCGG 151
Db 275 AAGTATATCATGCGCTGATGAGCGTTGCGCCACCCCGGACCGGACATGCGCTGATGGCGC 334
QY 152 CCATGGGCGATGCTGCGAGCGGGCGGCGCAACCCGGGTGCAAGCTGTCGGGCTGGTGG 211
Db 335 TCACCCCGGCTCTTCGCGCGCATGACCGGACACCGGCAACCTCGACCCGAGACAGTGG 394
QY 212 TCGCGCCCGGCTTCATCGACTGCGACACCGACGACACAACTACCTGCTAGGCGGCGG 271
Db 395 CCGCGCGGCTCACACCCCGGCACTGGGCGGTGCGGGCTCCACCTTGCGGGCGCGCCCT 454
QY 272 ACATGACGCCCAAGATCTCGCAGGGGCTCACACGAGGTGTCAGGGCAATTGGCGCATCA 331
Db 455 GCGCGCGCGACCAAGCTGCGAAGGTGCGGAGAGACAGGACGCGCTGCGGCTGTACTTCGACG 514
QY 332 GCTTGGCGCGCTGGCGCGACGCAACCGCGCGCGCGCGCTGGACCTGCTGGACGAAAGCG 391
Db 515 CCGGCGACGCGCTCGGCTGCGCGCGCTGACGAGCGCGCGCGCGCGCGCTGCGCGAGCGCG 574
QY 392 GCTTTCACGTTTTCGAGCGCTTCGCGCGCTGACGCGCTGGCGGCGCGCGCGCGCGCG 451
Db 575 AGGCTTCAGCTTCCA--CGCACCAAGGCGGTCAAGCGCTTCAGAGGCGCGCGCGCTCG 631
QY 452 CCGTCAAGCGCGCGCTGATGCTGAGGCGCTTCAGAGCTGCGCGCGCGGTCAATGCGGACT 511
Db 632 TCACCGAGAGCGCGGACCTCGCGCGCGGATCGCGCGCTCCACAACTTCGCGCTTCGACG 691
QY 512 TGCAGCG 571
Db 692 TGGCG 751
QY 572 TGGCG 631
Db 752 TGGCGCTGACCTCTCTGAGCGCGCTTCCCGAGTATCGACCGGACCGAGCGCGCGCGCG 811
QY 632 CCAACGAGAGATCATGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691
Db 812 CCGCGCTACCGCGAGACACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 871
QY 692 CCCACATGCGCGAGAGCGGAGCAGCATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
Db 872 ACGGCGCTCAACACCAACCACTAGCTGATCGTCAAGATCGACGAGCGCGCGCGCGCGCG 931
QY 752 GCGCGAGCTGAGAGTGGCGCGGTGATCTCGACACCAAGTCTATGGCGCGCGCGCGCGCG 811
Db 932 ACCGCGACCTGCTATGAGAGTCTCTGAAGGCGGAGGCGTGCACACCGCGCGCTACTTCT 991
QY 812 TCGCGCGCTGCGCGGAGAGCGCTGCGCGCTGATCGAG-----GCCGCGATGGCGCGCG 862
Db 992 CCGCGGCGCTGCGCGAGCTGAGACCGCTACCGCGCGGAGCGCGCGCGCGCGCGCGCGCG 1051
QY 863 AAGACGCTCGCGTGAAGCGGCTATCCTACGTGGCGCGCGCTCCACATGCTCAAGCAGAGCG 922
Db 1052 CCGAAGCGCTCGCGCGCGCGCGGTCTCTCTGCGCGAGCGCGCGCGCGCGCGCGCGCG 1111
QY 923 GCGTGTGCTGCG 982
Db 1112 ACATTCGCGCGGCTGCGCGAGCTGCTGCTGCTGCGCGAGACCGCGCGCGCGCGCGCGCG 1171
QY 983 GCGCGAGCTGAGTGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
Db 1172 CCGCGACCGCGAGACG 1204

GenCore version 5.1.4-P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 23:54:40 ; Search time 395 Seconds

(without alignments)
10022.818 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 1758
Sequence: 1 gaattccactgacgcgcgga.....ccctgacgcgcgaagctt 1758

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1758	100.0	1758	22	AAC91797
2	1440.6	81.9	1473	21	AAAI5140
3	103	5.9	114955	20	AAAX3491
4	101.2	5.8	312	24	ABN24790
5	99.2	5.6	65140	22	AAAD17188
6	99.2	5.4	123401	22	AAAD17186
7	94.2	5.4	8438	15	AAQ73500
8	93.8	5.3	114955	20	AAAX3491
9	93.4	5.3	27541	22	AAAD17185

10	86	4.9	88421	24	AAI40781	88421nt genomic DN
11	84.6	4.8	109519	22	AA508693	Micromonospora DNA
12	81.8	4.7	1248	21	AA287286	S. venezuelae deso
13	81.8	4.7	1248	21	AA287286	Streptomyces venez
14	81.8	4.7	12441	21	AA287284	S. venezuelae deso
15	81.8	4.7	13613	21	AA287319	Streptomyces venez
16	81.8	4.7	13613	24	AA287319	Streptomyces venez
17	79.8	4.5	1320	23	AA022481	Complete Mitocycin
18	78.2	4.4	3957	22	AA095866	Human herpesvirus
19	78.2	4.4	12249	21	AA055840	Human herpesvirus
20	78.2	4.4	18331	21	AA055857	Complete Mitocycin
21	78.2	4.4	154746	24	AA025519	Complete Mitocycin
22	78.2	4.4	154746	24	AA025519	Complete Mitocycin
23	76.2	4.3	5970	21	AA055635	Human herpesvirus
24	76.2	4.3	12001	16	AA076213	Nucleotide sequenc
25	76.2	4.3	12001	16	AA076213	Confir 002 from co
26	76.2	4.3	12001	16	AA076213	HSV 1/ST region.
27	76.2	4.3	12001	16	AA076213	Sortangium cellul
28	75.8	4.3	38857	17	AA058471	Nucleotide sequenc
29	75.8	4.3	4020	18	AA091361	Tyrosine synthase
30	75.6	4.3	43280	10	AA092408	Sequence encoding
31	75.4	4.3	2061	13	AA022482	groEL-1 gene codin
32	75.4	4.3	2061	13	AA022482	groEL-1 gene codin
33	75.4	4.3	2061	13	AA022482	groEL-1 gene codin
34	73.8	4.2	15872	18	AA022485	Streptomyces venez
35	73.8	4.2	15872	18	AA022485	Streptomyces venez
36	73.8	4.2	15872	18	AA022485	Streptomyces venez
37	73.4	4.2	4257	19	AA068520	The nucleotide seq
38	73.4	4.2	4257	19	AA068520	Infected cell prot
39	73.4	4.2	5224	22	AA090079	L05390 cDNA clone.
40	72.8	4.1	11604	22	AA032022	Human MERT1 relate
41	72.8	4.1	15079	16	AA091580	Streptomyces clav
42	72.8	4.1	15079	22	AA091580	S. clavuligerus cl
43	72.8	4.1	3189	23	AA51474	Streptomyces clavu
44	72.2	4.1	13842	21	AA287297	Pseudomonas aerugi
45	72.2	4.1	36778	21	AA287318	S. venezuelae macr

ALIGNMENTS

RESULT 1
AAC91797 standard; DNA; 1758 BP.

AC AAC91797:

27-MAR-2001 (first entry)

Alcaligenes xylosoxidans subspecies xylosoxidans D-aminoacylase DNA.

KW D-aminoacylase; zinc tolerant host; recombinant production;
KW zinc-enhanced expression; D-form amino acid synthesis;
KW antibiotic production; peptide drug; pharmaceutical manufacturing;
KW ds.

OS Alcaligenes xylosoxidans.

FN WO200078926-A1.

PD 28-DEC-2000.

PF 15-JUN-2000; 2000WO-JP03932.

PR 17-JUN-1999; 99JP-0170555.

PA (AMAN-) AMANO ENZYME INC.

PI Takeuchi K, Koide Y, Hirose Y, Moriyuchi M, Isobe K;
DR WPL: 2001-080828/09.
DR P-PSDB; AAB48975.

PT Transformed microorganism from zinc-tolerant host for selective
 PT production of D-aminoacylase, useful in synthesis of high
 PT optical purity D-amino-acids for antibiotic side-chains and peptide
 PT drugs -

PS Claim 2; Page 14-18; 22pp; Japanese.

CC The invention relates to a recombinant zinc-tolerant microorganism which
 CC expresses the D-aminoacylase from *Alcaligenes xylosoxidans* subspecies
 CC *xylosoxidans*. The presence of zinc ions in the culture medium enhances
 CC the process of recombinant D-aminoacylase gene, and the invention also relates to
 CC microorganism of the invention. The recombinant microorganism is used
 CC for the selective production of D-aminoacylase, which is useful in the
 CC synthesis of high optical purity D-form amino acids for use in the
 CC production of antibiotics and peptide drugs. The present sequence
 CC represents DNA encoding D-aminoacylase from *Alcaligenes xylosoxidans*
 CC subspecies *xylosoxidans*.

SO Sequence 1758 BP; 281 A; 662 C; 568 G; 247 T; 0 other;

Query Match 100.0%; Score 1758; DB 22; Length 1758;
 Best Local Similarity 100.0%; Pred. No. 5.8e-267;
 Matches 1758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCACCTTGATGCGGAGAGAGATTTCCATGTCACATCCGATTCCAGCCCTTC 60
 Db 1 GAATTCACCTTGATGCGGAGAGAGATTTCCATGTCACATCCGATTCCAGCCCTTC 60
 QY 61 GACCTGTCGTCGCGGCGGCGGACCTCATGACGAGCAGACACCCCGGCGCGCGC 120
 Db 61 GACCTGTCGTCGCGGCGGCGGACCTCATGACGAGCAGACACCCCGGCGCGCGC 120
 QY 121 GACCTGTCGTCGCGGCGGCGGACCTCATGACGAGCAGACACCCCGGCGCGCGC 180
 Db 121 GACCTGTCGTCGCGGCGGCGGACCTCATGACGAGCAGACACCCCGGCGCGCGC 180
 QY 181 CACACCGGCGTCGACGTCGTCGCGGCGGCGGACCTCATGACGAGCAGACACCC 240
 Db 181 CACACCGGCGTCGACGTCGTCGCGGCGGCGGACCTCATGACGAGCAGACACCC 240
 QY 241 CACGACGACAACTACTGTCGTCGAGCGTCGACGACGACGACGACGACGACGAC 300
 Db 241 CACGACGACAACTACTGTCGTCGAGCGTCGACGACGACGACGACGACGACGAC 300
 QY 301 ACCACGCTGTCACGCGGCAATTCGCGCATCAGCTGCGCGGTCGGCGGACGCG 360
 Db 301 ACCACGCTGTCACGCGGCAATTCGCGCATCAGCTGCGCGGTCGGCGGACGCG 360
 QY 361 CCGGCGGCGGTCGACCTGCTGACGAGAGCGGCTCTTACCGTTTCGAGCGCTTC 420
 Db 361 CCGGCGGCGGTCGACCTGCTGACGAGAGCGGCTCTTACCGTTTCGAGCGCTTC 420
 QY 421 TACCTGACGCGCTTGGCGGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
 Db 421 TACCTGACGCGCTTGGCGGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
 QY 481 TCAAGCTGCGCGCGGCTCATGCGGCTTGCAGCGCGGCGGCGGCGGCGGCGG 540
 Db 481 TCAAGCTGCGCGCGGCTCATGCGGCTTGCAGCGCGGCGGCGGCGGCGGCGG 540
 QY 541 GGGGCGCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
 Db 541 GGGGCGCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
 QY 601 GGGGCGCTTCTACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 Db 601 GGGGCGCTTCTACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 QY 661 CCGCTGAGCGCGCATGCGGCGGCTATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
 Db 661 CCGCTGAGCGCGCATGCGGCGGCTATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720

QY 721 GTGCGCGCGGTCGAGGAGAACTTCCGATGCGCGGCGGCGGCGGCGGCGGCGG 780
 Db 721 GTGCGCGCGGTCGAGGAGAACTTCCGATGCGCGGCGGCGGCGGCGGCGGCGG 780
 QY 781 TCGCACCACAAAGGTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
 Db 781 TCGCACCACAAAGGTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
 QY 841 ATGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
 Db 841 ATGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
 QY 901 TCCACATGCTCAAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
 Db 901 TCCACATGCTCAAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
 QY 961 AAGCCCTTCCCGAAGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
 Db 961 AAGCCCTTCCCGAAGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
 QY 1021 TCCAGTACGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 Db 1021 TCCAGTACGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 QY 1081 GAACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
 Db 1081 GAACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
 QY 1141 CCGCAGCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
 Db 1141 CCGCAGCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
 QY 1201 TATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
 Db 1201 TATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
 QY 1261 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
 Db 1261 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
 QY 1321 CTGCTGCTGTCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
 Db 1321 CTGCTGCTGTCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
 QY 1381 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
 Db 1381 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
 QY 1441 TTCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
 Db 1441 TTCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
 QY 1501 CCTTCAATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
 Db 1501 CCTTCAATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
 QY 1561 CGCTACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
 Db 1561 CGCTACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
 QY 1621 GTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
 Db 1621 GTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
 QY 1681 TCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740
 Db 1681 TCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740
 QY 1741 CTGAGCTTCAAGAGGCTT 1758
 Db 1741 CTGAGCTTCAAGAGGCTT 1758

RESULT 2

AA15140

ID AA15140 standard; DNA; 1473 BP.

AC AA15140;

XX 21-AUG-2000 (first entry)

DE DNA encoding an aminocyclase enzyme.

XX Aminocyclase; N-acetyl-D-tryptophan; R-N-acetyl-2-thienylalanine;
 KM R-N-acetyl-4-chlorophenylalanine; D-amino acid; N-acyl amino acid;
 KW pesticide; antibiotic; ss.

XX Alkaligenes sp.

XX Key Location/Qualifiers
 FT CDS 7..1461
 FT /tag- a
 FT /product- "aminocyclase"

XX MO200023598-A1.

XX 27-APR-2000.

XX 20-OCT-1999; 99MO-GB03458.

XX 20-OCT-1998; 98GB-0022947.

XX 01-APR-1999; 99GB-0007739.

XX (CHIR-) CHIROTECH TECHNOLOGY LTD.

XX Taylor StC, Brown RC;

XX WPI; 2000-339697/29.

XX P-PSDB; AAY84943.

XX New isolated D-amino acylase enzyme, useful for the preparation of
 PT D-amino acids for use as intermediates in the preparation of
 PT pesticides, antibiotics and other pharmaceuticals

XX Claim 7: Page 18-20; 26pp; English.

XX The present sequence encodes an aminocyclase enzyme. The enzyme is
 CC capable of hydrolysing N-acetyl-D-tryptophan at a substrate
 CC concentration of 10 grams per litre, and exhibiting faster conversion
 CC of R-N-acetyl-2-thienylalanine than of
 CC R-N-acetyl-4-chlorophenylalanine. The enzyme can be used for the
 CC preparation of a D-amino acid by conversion of a corresponding
 CC D-N-acyl amino acid. It can also be used for resolving a racemic
 CC mixture of N-acyl amino acids and deprotecting optically-enriched
 CC N-acyl amino acids. The D-amino acids are useful as intermediates in
 CC the production of various pesticides, antibiotics and other
 CC pharmaceuticals.

XX Sequence 1473 BP; 227 A; 559 C; 481 G; 206 T; 0 other;

XX Query Match 81.9%; Score 1440.6; DB 21; Length 1473;

XX Best Local Similarity 99.0%; Pred. No. 2.9e-217;

XX Matches 1449; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 31 TCCATGTCGCAATCGATGCCAGCCCTTCGACCTGCTGCGGGGCGGACCTTCATC 90
 DB 4 TCCATGTCGCAATCGATGCCAGCCCTTCGACCTGCTGCGGGGCGGACCTTCATC 63
 QY 91 GACGCGACCAACACCCGCGGGGCGGCGGACGACCTGCGGGGCGGACCGATCGCC 150
 DB 64 GACGCGACCAACACCCGCGGGGCGGCGGACGACCTGCGGGGCGGACCGATCGCC 123
 QY 151 GCCATCGGCGATCTGTGCGGAGCGCGCGGACCAACCGGGGTGACGCTGCGGGCTGGTG 210
 DB 124 GCCATCGGCGATCTGTGCGGAGCGCGCGGACCAACCGGGGTGACGCTGCGGGCTGGTG 183

QY 211 GTCCGCGCCGCGCTTCATCGACTGCGACACCGACGACCACTACCTGCTGAGCGTGC 270
 DB 184 GTCCGCGCCGCGCTTCATCGACTGCGACACCGACGACCACTACCTGCTGAGCGCGC 243
 QY 271 GACATGACGCCCAAGATCTGCGAGGGGTCAACACGGGTGTACAGGGGCAATTTGGGCATC 330
 DB 244 GACATGACGCCCAAGATCTGCGAGGGGTCAACACGGGTGTACAGGGGCAATTTGGGCATC 303
 QY 331 AGCGTGGCGCGCGCTGCGGCGACGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
 DB 304 AGCGTGGCGCGCGCTGCGGCGACGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 363
 QY 391 GCGCTTACGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCG 450
 DB 364 GCGTGTACCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCG 423
 QY 451 GCGGTCAACGCGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCG 510
 DB 424 GCGGTCAACGCGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCG 483
 QY 511 TTGAGCG 570
 DB 484 TTGAGCG 543
 QY 571 ATGGCCAGCG 630
 DB 544 ATGGCCAGCG 603
 QY 631 ACCACCGAAGATCTATCGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690
 DB 604 ACCACCGAAGATCTATCGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 663
 QY 691 ACCACATGCG 750
 DB 664 ACCACATGCG 723
 QY 751 GCGCGCGAGCTGAGCGTGGCGGTGTGATCTGCAACCAAGGTGAGGCGCGCGCGCGCG 810
 DB 724 GCGCGCGAGCTGAGCGTGGCGGTGTGATCTGCAACCAAGGTGAGGCGCGCGCGCGCG 783
 QY 811 TTGCGCGCGCTGCGGAGGAGCGCTGCGGTGTGATCTGCAACCAAGGTGAGGCGCGCGCG 870
 DB 784 TTGCGCGCGCTGCGGAGGAGCGCTGCGGTGTGATCTGCAACCAAGGTGAGGCGCGCGCG 843
 QY 871 TTGCGGAGCGCTATCTGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 930
 DB 844 TTGCGGAGCGCTATCTGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
 QY 931 CTGCGCGGAGCG 990
 DB 904 CTGCGCGGAGCG 963
 QY 991 CTGATGAAGTCTGCG 1050
 DB 964 CTGATGAAGTCTGCG 1023
 QY 1051 CCG 1110
 DB 1024 CCG 1083
 QY 1111 GCGCGCGAGCAATGATGCGGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1170
 DB 1084 GCGCGCGAGCAATGATGCGGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1143
 QY 1171 TGGGCGCGCTTCCGCGCGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1230
 DB 1144 TGGGCGCGCTTCCGCGCGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1203
 QY 1231 GAGACGCGGTATGGAAGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1290
 DB 1204 GAGACGCGGTATGGAAGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1263
 QY 1291 GGGGAGCTGAGCG 1350

RESULT 4	ABN24790	ABN24790 standard; cDNA; 312 BP.
ID	ABN24790	
AC	ABN24790;	
XX		
DT	24-JUN-2002	(first entry)
XX		
DE	Human ORFX polynucleotide sequence SEQ ID NO:18057.	
XX		
KW	Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;	
KW	hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;	
KW	degenerative disorder; osteoarthritis; neurodegenerative disorder;	
KW	cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;	
KW	hypertension; hypothyroidism; cholesterol ester storage disease;	
KW	immune deficiency; immune disorder; infectious disease;	
KW	autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;	
KW	myasthenia gravis; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	MO200192523-A2.	
XX		
DD	06-DEC-2001.	

The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-1149) (see Table 1 in the specification). ABN15762 to ABN27252 encode the human ORFX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with an ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, and periodontal disease, and for gut protection or regeneration or peritoneal disease, and for gut repulsion injury in various tissues and conditions resulting from systemic cytokine damage.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Query Match	5.8%;	Score 101.2;	DB 24;	Length 312;
Best Local Similarity	58.5%;	Pred. No. 1.6e-07;		
Matches 175; Conservative	0;	Mismatches 124;	Indels 0;	Gaps 0

629 CCACCAACGAGAAATCATCTAGAGTGTGCCGGCTGAGGCCGATCTTACG 688
 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2 CGACCAACGAGAAATGATGGCGCTGGCCGAGACGGCTGGCCAGACTGGCCGATACGA 61
 689 CCACCCATCTCCGCGAGCAAGAGCGAGACATCGTGGCCGCGCTGGAGGAAACTTTCGCA 748
 62 CGACCAATCTCCGCGAGCAAGAAATTTGGCGCGATCTCTCATGTCGATGGACGAGGGGATACCGCG 122
 749 TCGGCGCGGAGACTGGAGAGTCCCGGTGTATCTCGACCAAGGTCAATGGCCACGCCA 808
 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 122 TCGGCAAAACATCCGCGCGTCCCGGTGTATCTCAACACTGAAATGCCACGGCCCTGCCA 181
 809 ATTTCGCGCGCTCGCGGAGACGCTCCGCTATATCGAGGCGCCCATCGCGCGCAGAGAGC 868
 182 ATTGGGCGCGCGCGGAGAGGTCTCGAATTCATGGAAGGTGCACAGTCTATATGACCGG 241
 869 TCTGTGAGAGCGGTATTCCTACGTGGCCGCTCCACCATGCTCAAGCAGGACCGCGTG 927
 242 TGGGTGCGATGCTATCTCTCAACCGCAAGGTGCTCCACACTGAGACTCAAGCAGGTG 300

PA (SIN/ - SINVENT AS.
PA (DZIE/ DZIEGLEWSKA H.
PA (ZOTC/ ZOTCHEV S B.
PA (SEKU/ SEKUROVA O N.
PA (EJAE/ FJAEVRY E.
PA (BRAU/ BRAUTASET T.
PA (STRO/ .STROM A R.
XX
XX
PI Zotchev SB, Sekurova ON, Fjaerylk E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX
DR MPI: 2001-557614/62.
DR P-PDSB, AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137,
XX
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
PT
PT
PS
PS Claim 2; Page 116-151; 266pp; English.
XX
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nyl1 DNA of nystatin PKS gene cluster
XX
XX Sequence 65140 BP: 8270 A: 25171 C: 3373 G: 940 C: 1400 T: 1400

Oy 62 ACCTGCTGCTGCGCGGCGGACCCCTCATCGACGGCAGCAACACCCCGGGGGCGCGCG 121
 Db 19698 ACGAGCGCGCTGCGCGGCACTGCGCTTCCAGTACGGCCCGCGCTTCCAGGGCGCTGCGGGCG 19757
 Oy 122 ACCTGGCGCTGCGGGGAGCGCGCATGCGCGCCCAATCGCGCATCTGTGCGAGCGCGCGCG 181
 Db 19758 CCTGGCGCGCGACACCGAGGTGTACGCGCGGAACTGGCCCTGTGCCGACGGCGCGGACACG 19817
 Oy 182 ACGACCGGGTTCGACGTGTGCGGCGCTGTGTGCGCGCGCGCTTCATTCGACTCGCACCC 241
 Db 19818 ACCCGCGCGCTTTCGGACTGCACCCCGGCGCTGTGTGAGCGCCGACAAACAGCGCGCGCT 19877
 Oy 242 ACGAGCAACAATCTGCTGCAAGCGCGCGGAGCATGACGCCCAAGATCTTCGAGGGCGCTCA 301
 Db 19878 ACGCGACCTCGGCGCGCATACGCGGGGCGCTTCCGCTTGTGCGGGAGGCGGCTGCG 19937
 Oy 302 CCACGGTGTACAGGGCAATTGCGGCATGACGCTGGCGCGCGTGGCGGACGCCCAACCGC 361
 Db 19938 TCGCGCGCGCGCGCGCCACACACCTGTGCGCGCGGATGCGCCCGCGCGGCGAGGACACG 19997
 Oy 362 CGCGCGCGCTGGACCTGCTGAGCAAGCGCGCTCTTACGTTTCGAGCGCTTCGCGCACT 421
 Db 19998 TCACCATGCGCGCTTACGACAGCGCGCGCGCACCTGTCTGCTGCGTCACTCCCTGTCT 20057
 Oy 422 ACCTGAGCGCTGCGGGGCGCACCGCGGCGCGCTCAAGCGCGCGCTGTATGTGGCCATT 481
 Db 20058 CCGCGAGGTCCGCGCGACGACCGCGCGCGCGCGCACCGTCAACCGGACTCCCTCT 20117
 Oy 482 CAAGCGCTGCGCGCGGTATGCGCGACTTGCAGCGCGCGCGCACCGAGAGGAATGCG 541
 Db 20118 TCCAGTGTGATGGAACCCCGCTCCAGGGCGCGCGCGCGCGCGCACCGGCAACGTTGGCG 20177
 Oy 542 CGGCGCATGCGGAGCTGTGCGCGAGAGACCAATGGCGACGCGCGCGCATGCGCATTTGACCG 601
 Db 20178 TCTCTGGCGCGGACCGGAGCGCGCTGCGGACACCTTCGCGCGCACCGGCAATCGGACCA 20237
 Oy 602 GCGGCTTTCACCGCGCGCGCGCGCGCGCACACCGAGAGATTCATGAGGTGTGCGCGC 661
 Db 20238 CCGCGCGCGCGGACCTGCGCGCGCTTGTGCGGAGCGCGGAAAGGCGCGTCCCGGACCTGTGCG 20297
 Oy 662 CGGTGACCGGCACTGGCGCGCATCTACGCCACCCACATGCGGAGCGAGGCGACATGCG 721

Db 20298 TCACACCTCTACACACACACCCCGGCGCCCGCTCCCGACGCGCGACCGCACCG 20357
 QY 722 TGGCCGCGCTGAGAGAAACCTTCGCAATCGGCGGAGCTGAGCTGCGGTGATCT 781
 Db 20358 CCGCGCTCTCTGCGCTCCGCAACAGTGGCTGGCGAGACCGCTTCCGACGCGCCG 20417
 QY 782 CGACACCAAGATGATGAGCCGACCCCAATTTCGGCGCTCGCGGAGACGCTGCCGTGA 841
 Db 20418 TGTCTCTGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20477
 QY 842 TCGAGCG 20477
 Db 20478 GCTGATGCG 20537
 QY 902 CCACATGCTTAACGAGACCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 961
 Db 20538 CCGACACCG 20591
 QY 962 AGCCCTTCCCGACCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1021
 Db 20592 ACAGACGAGCG 20651
 QY 1022 CCAAGTACGAGCTGTGCTCCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1081
 Db 20652 TCCCGCTCG 20711
 QY 1082 AACCGGACGTCGAGCGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1141
 Db 20712 GCAACCG 20768
 QY 1142 CGACGAGCG 1201
 Db 20769 GCCACGCTGCTGCG 20828
 QY 1202 ATGCG 1261
 Db 20829 CCGACACTACCG 20888
 QY 1262 CCG 1321
 Db 20889 CCGC-----CTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20939
 QY 1322 TGTGCTGTTGACCG 1381
 Db 20940 TGTGCTGTTGACCG 20999
 QY 1382 GCG 1441
 Db 21000 GCTGAGACACCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21059
 QY 1442 TCACCG 1501
 Db 21060 GCGACCTGACCTGAGCGCGCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 21119
 QY 1502 CCTTAC-----AATCCGCGGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1555
 Db 21120 CCGGACAGGCAATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21179
 QY 1556 CAACCGCTGATGCG 1615
 Db 21180 CCGGACAGGCG 21239
 QY 1616 AGGAAGTATGCG 1672
 Db 21240 TGACAGAGCG 21299
 QY 1673 AAGACCTGTCGAGCG 1716
 Db 21300 TGACCTGTAACAGGCG 21343

RESULT 6

AAD17186
 ID AAD17186 standard; DNA: 125401 BP.
 AC AAD17186;
 DT 29-NOV-2001 (first entry)
 DE Streptomycetes noursei nystatin PKS gene cluster DNA.
 KW polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 KW antifungal; antibiotic; ds.
 OS Streptomycetes noursei.
 XX
 FH Key location/qualifiers
 FT CDS 6337..34771
 FT /tag- a
 FT /product- "NysI complete protein"
 FT CDS 34792..51099
 FT /tag- b
 FT /product- "NysJ protein"
 FT CDS 51155..57355
 FT /tag- c
 FT /product- "NysK protein"
 FT CDS 57503..58687
 FT /tag- d
 FT /product- "NysL protein"
 FT CDS complement (58786..58980)
 FT /tag- e
 FT /product- "NysM protein"
 FT CDS complement (59045..60241)
 FT /tag- f
 FT /product- "NysN protein"
 FT CDS complement (60238..61296)
 FT /tag- g
 FT /product- "NysD2 complete protein"
 FT CDS 120628..121308
 FT /tag- h
 FT /product- "NysR4 (long) protein"
 XX
 PN WO200159126-A2.
 XX
 PD 16-AUG-2001.
 XX
 PF 08-FEB-2001; 2001WO-GB00509.
 XX
 PR 08-FEB-2000; 2000GB-0002840.
 PR 10-APR-2000; 2000GB-0008786.
 PR 14-APR-2000; 2000GB-0009387.
 XX
 PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
 PA (SNTF) SINTER SMITTELESEN IND TEK FORSK.
 PA (ALPH-) ALPHARMA AS.
 PA (STNV-) STINVENT AS.
 PA (DZIE/) DZIELEWSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (FJAE/) FJAEVYK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 XX
 PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX
 DR MPI: 2001-557614/62.
 DR P-PSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
 DR AAE10149, AAE10150.
 XX
 PT New nystatin polyketide synthase polynucleotides and polypeptides,
 PT useful as antibiotics and antifungals -

D	5427	GC	CGCGGGCCCCGGCGAGAACGCCGCTCCACCAACCAACACTCTTCGCTCATGCGCGAC	5368
Q	271	GACATGACGCCCAAGATCTCGAGGGCGTCAACCAAGGTGTACGAGGCAATGGCGATC	330	
D	5367	GGGCCCCCGGACCGAGGAGGGGCGCGCTGTCTACCCCGCTCGGGAGGCCCTGGCCCGGGCTCC	5308	
Q	331	AGCGTGGGGCGGCTGGGGCCAGGCCAACCCCGCGCCCTCTGGACTGTCTTGACAGAAAGC	390	
D	5307	GACCTTCGGGCGGACAGGGCGCGCTCCGCTACAGGGCGGCGCGGAGCTCCCGGAGGGGCTC	5248	
Q	391	GGCTTTACCGCTTTGAGCGGCTTCGCGACTACCTGGACGCGGTTCGGGGACCGCCGGCG	450	
D	5247	TGGAGCAGGAGCAGAGTCCGCGACAGCGCGCGCCCGCCGCTACCGCGCCGCGGCGCTGTG	5188	
Q	451	GCGCTCAACGCCCGCTGTATGTGTGGCCATTCAACGCTGCGCGCGCGGTCACTCCGAC	510	
D	5187	CCGCTCTTCATCCCGGAGATGGGGGACTCGAGAAAGCAGCAGGCGCTTCGTGCGCTC	5128	
Q	511	TTGAGAGCGCGCGCGACCGAGAGAGAAATGCGCGCCATGCGGGGACTGTGGCCAGAAAGC	570	
D	5127	ATCTACAGCGGGGCGCGCGCGCGAGGCCATGTCTGTGCTGCAAAACCGCGAGATCGAGCC	5068	
Q	571	ATGGCCAGCGGCGCCATCGGCAATTCAGCGCGCTTTCACCGCGCCCGCGCGCGCGCC	630	
D	5067	CCGACACGAGGCTTCAACCAAGTTCTGCGACAGCGCGGGTCCACGCGCGCCCAAGGCAAGC	5008	
Q	631	ACCAACGAGAGATCATCGAGGTGTGCGCGCGCGCTGAGCGCGCATGGCGGCATCTACGC	690	
D	5007	TCTCTCATACCGGACGAGCTGACCCCGCGCTGCGGACATCGGGAGCGCATGTGGCGCC	4948	
Q	691	ACCCACATGGCGCAGAGAGGCGAGCACATGTGAGCGCGCTGAGAGAAACTTCGCGATC	750	
D	4947	CAGAGACCGGCTGTGGCGCTTCGCGACGCGGTGAGCGCGGTGGCATGAGCCCGCTAC	4888	
Q	751	G---GCGGAGGTGAGCGTGGCGGTATCTCGACCAAGATCATGGGCGAGCC	807	
D	4887	GATCGCACCCAGAAAGACTTTCATCTGCAAGAGCTTCGCGGGGCTACGCGCATGTGGCC	4828	
Q	808	AATTTGGCGCTTCGCGGAGAGAGCGTGCCTCATGAGGCGCGCATGGGCGCGCAGAGAC	867	
D	4827	TACCGGGCGCGCGCGGAGACCCCGCGCGCGGGAGAGGACCTTCGAGGCCCTCTGCGCC	4768	
Q	868	GTCCTGCTGAGCGCGTATCTCTACG---TGCGCGCTCCACCATGCTCAACGAGAGCCG	924	
D	4767	CGGTCGCGCGCGCTTCGCGCGCGCGGACGCGCGCGGTGCGCCGCGAGCTGGCGGAC	4708	
Q	925	GTCCTGCTGAGCGGAGCAGCATCATCACTGCTGCAACGCTTCGCCCATGAGAGCGGG	984	
D	4707	GCTGTGCTCTGCGCTGCGCGCGCGCTGCTGACCGCTGTGCTCTGCGCGCTCGCGCTG	4648	
Q	985	CGGACCTGGAATGAAGTTCGCGCGCGGCGGCAATTCGAAGTACAGCTGTGGCCGAG	1044	
D	4647	CCCGGCGCGCGCGCGCGCGCGCGCGCGCTTCGGGCGCGCTGCGCTGAGGAGTGAACGCG	4588	
Q	1045	CTGAGCGCGCGCGCGCGCCATCTACTTCACTGATGAGAGAAACCGAGGTGACGCACTCG	1104	
D	4587	GCGGCTCTGCGCGCTCCGCGACCGAGATCCCGGGGCGCGCGCGAGGCGAGCAGAGCC	4528	
Q	1105	GCGTTTCGCGCGCACATGAATGATGAGTTCGAGCGCTGCGCGCAGAGAGCGCGGATCG	1164	
D	4527	GCGGACTGTGTGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	4475	
Q	1165	GCGCTGTGGGCACTTCCCGCGGGGTGCTGGGGCACTATGCGCGGACCTTGCGGCTGTTC	1224	
D	4474	CGTGAAGCGGGCGCGCGCGCGG-----AGGCGCGCTGAGACTTACGCGCGCGCTTTC	4420	
Q	1225	CGGCTGAGAGCGCGGCTATGGAAGATGACCGGCTGACCGCGCGCGCGCTGTGGCGCGC	1284	
D	4419	GCGCGCGCGCACTGTGGCGCGGGCGCGCGCTGCGCGAGCGCGCGCGCGCGCGCGCGCG	4360	
Q	1285	GGGCGCGGAGCTGACAGCGCGGATCTTCGCGGACTGTGTGTTTGACCCGCGCACG	1344	
D	4359	GAGCGCGCGCGCGCGCTTCCCGCGGCTGTGGCGCGGAGAGACCGCGCTGTGTATGCT	4300	

OY 1345 GTGGCCGATACCCGACCTTCGACACCTTCAGAGCGCGCGCGG 1391
 Db 4299 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 4253

RESULT 8

AXX53491 standard; DNA; 114955 BP.

AXX53491;
 05-JUL-1999 (first entry)

Human adenosine A1 receptor antisense oligonucleotide fragment.

Antisense oligonucleotide; multiple target; antisense treatment;
 impaired respiration; inflammation; lung disease;
 pulmonary vasoconstriction; inflammation; allergic rhinitis;
 acute asthma; allergy; asthma; impaired respiration;
 respiratory distress syndrome; pain; cystic fibrosis;
 pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 colon cancer; breast cancer; lung cancer; pancreatic cancer;
 hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 prostate cancer; ss.

Synthetic.

MO9913886-A1.

25-MAR-1999.

17-SEP-1998; 98WO-US19419.

09-JUN-1998; 98US-0093972.

17-SEP-1997; 97US-0059160.

(TREC-) UNIV EAST CAROLINA.

Nyce JW;

WPI; 1999-229400/19.

New antisense oligonucleotides used in treatment of, e.g. pulmonary
 vasoconstriction

Disclosure; Page 37; 120pp; English.

The specification describes antisense oligonucleotides (AXX52869-X55271)
 directed against at least 2 mRNAs selected from target genes, coding and
 non-coding regions of RNAs corresponding to target genes, coding and
 initiation codons, genomic flanking regions, intron-exon borders, the
 5'-end, the 3'-end and the juxta-section between coding and non-coding
 regions and all segments of RNAs encoding proteins associated with one
 or more diseases, conditions or mixtures. The antisense oligonucleotides
 may be derived from sequences AXX5272-74. These multiple target
 oligonucleotides (specifically AXX5272-74) can be used for the
 antisense treatment of diseases and conditions. Typical diseases and
 conditions are those associated with impaired respiration and
 inflammation, including lung diseases, pulmonary vasoconstriction,
 inflammation, allergic rhinitis, acute asthma, allergies, asthma, impaired
 respiration, respiratory distress syndrome, pain, cystic fibrosis,
 pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
 obstructive pulmonary disease (COPD), and cancers such as leukemias,
 lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
 pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
 hepatic metastases, as well as all types of cancers which may metastasize
 or have metastasized to the lungs, including breast and prostate cancer.

Query Match

5.3%; Score 93.8; DB 20; Length 114955;

Best Local Similarity 32.6%; Pred. No. 1.2e-06;
 Matches 500; Conservative 131; Mismatches 880; Indels 25; Gaps 5;

OY 44 CGATTCCAGCCCTTCGACCTGCTGCGCGGCGGACCTTCATCGACGACGACCA 103
 Db 104365 CCGCGCGCGCGGGSNNNDNNCCGCGCGCGCGCGCGCGCGCGGCGGCGG 104424
 OY 104 CCGG 163
 Db 104425 GCGG 104484
 OY 164 TGTGAGACCGG 104484
 Db 104485 NCGG 104544
 OY 224 TCATGACTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 283
 Db 104545 NDNNCCGG 104604
 OY 284 AGATCTC--GCAGGCGCTCACCGGTGTCACGGGCAATTCGCGCATCAGCGCGCG 341
 Db 104605 DNNCCGG 104664
 OY 342 GCTGCGCGACCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 401
 Db 104665 GCGGCGG 104723
 OY 402 TTTGAGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 461
 Db 104724 GGGCGCGGNNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 104783
 OY 462 GCGCTGATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 521
 Db 104784 CCGG 104843
 OY 522 CGCCACCGCGAGGAAATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 581
 Db 104844 CCGG 104903
 OY 582 CGCCATCGCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 641
 Db 104904 GCGG 104963
 OY 642 GATCATCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 701
 Db 104964 NNDNNCCGG 105023
 OY 702 CGACGAAGCGGAGCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 761
 Db 105024 CCGCGCGCGGNNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 105073
 OY 762 GAGCTGCGCGGTGTGATCTGCGACACCAAGTCATGCGCGCGCGCGCGCGG 821
 Db 105074 GCGG 105133
 OY 822 GCGCGAGCGG 881
 Db 105134 NNDNNCCGG 105193
 OY 882 GTATCCCTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 941
 Db 105194 NNNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 105253
 OY 942 CACCATATCATCTGTGTGCAAGCCCTTCGCGCGCGCGCGCGCGCGCGCGG 1001
 Db 105254 DNNCCGG 105313
 OY 1002 CGG 1061
 Db 105314 CCGG 105373
 OY 1062 CATCTACTGATGATGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGG 1121

OY 559 GCCGAGGAAGCCATGCGGCGGCGCATGGCAATTGACCGGCGCTTCTACCGGCC 618
 DB 24029 ACCGTCGGCGTGAAGGTATACCAACACCGCGCGCGCTGGCGCGGATACCGAA 24088
 OY 619 GCCGCGCGCGCCACACCGAAGATCATCGAGGTGTGCGCGCGCTGAGCGCGATGCG 678
 DB 24089 GCCGCGCGCGCGACACAGGCGCGCTTCAAGCGGTGTGCTGCCGTGCGCGCTGCGACCGGC 24148
 OY 679 GGCACTACGCGCACCCACATCGCGGAGAGAGAGACATCGTGGCGCGCGCTGAGGAA 738
 DB 24149 GACCGCGCGCACCGCGGCGCGCGCGCTTCAACCGCTACACCGACCGCGCTGCGAGGCC 24208
 OY 739 ACCTTCGCGATGCGCGCGCGCGCTGAGCGCGCGGTGTGATCTGCAACACAGATGATG 798
 DB 24209 CTCGCGCGACGCGCGCGCGCGCGCGCTGTGAGACGTCAACCGCGCGCGCGCTG 24268
 OY 799 GGCGAGCGCGCATTTGCGCGCGCTGCGCGAGACGCTGCGCGGTGATGAGCGCGCGATGCG 858
 DB 24269 GGCGCGCGCGCGAAGGTACCGCGCGCGCGA-----CAGCGCGCGGTGTGCGCGCTG 24319
 OY 859 CGCGAGGACGCTGCTGCGAGCGGTATCCCTACGTGCGCGCGCTTCCACATGCTCAAGCAG 918
 DB 24320 GGCGCGCGCGCTGCGCGCTGAGACGTGCGCGCGCGGTTCGCGCGCGCGCTGCGCGCG 24379
 OY 919 GACCGCGGTGCTGCGCGCGCGCGCGCGCATCATGACGTGTCAGCGCGCTTCCCGGAACG 978
 DB 24380 ACCCTGAGCGCGCGCGCGCGCGCGCGCGCGCGGTGCGCGCGGTGCTGCGCGCGCTAC 24439
 OY 979 AGCG 1038
 DB 24440 GACCG 24496
 OY 1039 CCGGAGTGCAGCG 1098
 DB 24497 GCCG 24556
 OY 1099 ATCTGCGCGGTGCG 1158
 DB 24557 GCG 24616
 OY 1159 CATCGCGCGCGCTGT---GGGCGACCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1215
 DB 24617 CACTGCTGTGTACAGCG 24676
 OY 1216 GCGCTTTCG 1275
 DB 24677 GAACGTGAGGAACCTGCG 24736
 OY 1276 GCGCTGCG 1335
 DB 24737 GCGCTGCG 24796
 OY 1336 CG 1392
 DB 24797 ACCG 24856
 OY 1393 ATCCATTGCGGTACGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1452
 DB 24857 GTGCTGCG 24916
 OY 1453 CATGCG 1512
 DB 24917 CTCG 24973
 OY 1513 GCGGTGAAGCG 1563
 DB 24974 GGCACACTAGCG 25024

RESULT 10

AA140781

ID AA140781 standard; DNA: 88421 BP.

XX

AC	AA140781;	
XX		
DE	03-OCT-2002 (first entry)	
XX		
DE	88421nt genomic DNA of ramoplanin producing Actinoplanes sp.	
XX		
KW	Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;	
KW	biosynthesis gene cluster; bioengineering; peptide synthetase module;	
KW	adenylation domain; hydroxyphenylglycine; Hpg; antibiotic precursor;	
KW	chlorinate; lipopeptide; gene; ds.	
XX		
OS	Actinoplanes sp.	
XX		
Key	Location/Qualifiers	
FT	2077..3078	
FT	CDs	
FT	/*tag- a "Protein of ORF 1"	
FT	/product- "Protein of ORF 1"	
FT	3118..4032	
FT	/*tag- b "Protein of ORF 2"	
FT	/product- "Protein of ORF 2"	
FT	4038..5048	
FT	/*tag- c "Protein of ORF 3"	
FT	/product- "Protein of ORF 3"	
FT	complement (6665..5814)	
FT	/*tag- d "Protein of ORF 4"	
FT	/product- "Protein of ORF 4"	
FT	complement (7703..6693)	
FT	/*tag- e "Protein of ORF 5"	
FT	/product- "Protein of ORF 5"	
FT	complement (9464..8130)	
FT	/*tag- f "Protein of ORF 6"	
FT	/product- "Protein of ORF 6"	
FT	9691..10761	
FT	/*tag- g "Protein of ORF 7"	
FT	/product- "Protein of ORF 7"	
FT	complement (12751..10829)	
FT	/*tag- h "Protein of ORF 8"	
FT	/product- "Protein of ORF 8"	
FT	complement (13617..12802)	
FT	/*tag- i "Protein of ORF 9"	
FT	/product- "Protein of ORF 9"	
FT	complement (15203..13614)	
FT	/*tag- j "Protein of ORF 10"	
FT	/product- "Protein of ORF 10"	
FT	15591..15863	
FT	/*tag- k "Protein of ORF 11"	
FT	/product- "Protein of ORF 11"	
FT	15880..19035	
FT	/*tag- l "Protein of ORF 12"	
FT	/product- "Protein of ORF 12"	
FT	19032..39713	
FT	/*tag- m "Protein of ORF 13"	
FT	/product- "Protein of ORF 13"	
FT	39713..65800	
FT	/*tag- n "Protein of ORF 14"	
FT	/product- "Protein of ORF 14"	
FT	65826..66530	
FT	/*tag- o "Protein of ORF 15"	
FT	/product- "Protein of ORF 15"	
FT	66546..67370	
FT	/*tag- p "Protein of ORF 16"	
FT	/product- "Protein of ORF 16"	
FT	67384..70059	
FT	/*tag- q "Protein of ORF 17"	
FT	/product- "Protein of ORF 17"	
FT	70059..70662	
FT	/*tag- r "Protein of ORF 18"	
FT	/product- "Protein of ORF 18"	
FT	70659..71906	
FT	/*tag- s "Protein of ORF 19"	
FT	/product- "Protein of ORF 19"	
FT	complement (73439..71964)	
FT	/*tag- t	
FT	CDs	

FT CDS /product- "Protein of ORF 20"
complement (74216..73563).
/tag- u
FT CDS /product- "Protein of ORF 21"
complement (75424..74213)
/tag- v
FT CDS /product- "Protein of ORF 22"
75353..76464
/tag- w
FT CDS /product- "Protein of ORF 23"
complement (78110..76449)
/tag- x
FT CDS /product- "Protein of ORF 24"
complement (79864..78107)
/tag- y
FT CDS /product- "Protein of ORF 25"
complement (81624..79861)
/tag- z
FT CDS /product- "Protein of ORF 26"
complement (81909..81682)
/tag- aa
FT CDS /product- "Protein of ORF 27"
complement (82346..82062)
/tag- ab
FT CDS /product- "Protein of ORF 28"
82587..84446
/tag- ac
FT CDS /product- "Protein of ORF 29"
84481..85548
/tag- ad
FT CDS /product- "Protein of ORF 30"
85556..86845
/tag- ae
FT CDS /product- "Protein of ORF 31"
87372..86803
/tag- af
FT CDS /product- "Protein of ORF 32"
87494..88420
/tag- ag
/product- "Protein of ORF 33"
WO200231155-A2.
PN 18-APR-2002.
XX 15-OCT-2001; 2001WO-CA01462.
XX 13-OCT-2000; 2000US-239924P.
PR 12-APR-2001; 2001US-283296P.
PR 24-JUL-2001; 2001US-0910813.
XX (ECOP-) ECOPIN BIOSCIENCES INC.
PA Farnet CM, Zazopoulos E, Staffa A;
XX
XX WPI: 2002-435445/46
DR P-RSD: AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,
DR AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,
DR AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,
DR AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,
DR AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.
XX
XX Novel isolated ramoplanin biosynthetic pathway polypeptide useful for
PT chemically modifying biological molecule that is a substrate for a
PT polypeptide encoded by a ramoplanin biosynthesis gene cluster -
XX
XX Disclosure; Page 87-135; 212pp; English.

CC molecule with the isolated polypeptide, where the polypeptide chemically
CC modifies the biological molecule. The method comprises contacting the
CC biological molecule with at least two different polypeptides encoded by
CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the
CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated
CC gene cluster comprising the ORFs is useful as a substrate for
CC bioengineering of antibiotic structures. An isolated polypeptide or its
CC encoding nucleic acid sequence is useful for generating derivatives of
CC ramoplanin, for improving production or for producing variants of other
CC antibiotics of the peptide class. The isolated polypeptides are useful
CC for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain
CC in conjunction with other peptide synthetase modules and allowing the
CC incorporation of Thr into a peptide antibiotic precursor, for modifying
CC fatty acid structure and/or enhancing fatty acid incorporation into the
CC peptide antibiotic structure, for production of an hydroxyphenylglycine
CC (HPG)-containing peptide antibiotic, for enhancing secretion of
CC ramoplanin or its variants and derivatives, for enhancing uptake of
CC precursors for ramoplanin biosynthesis, for enhancing production of
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of
CC a peptide antibiotic precursor, and for designing specific nucleotide
CC probes and primers for identifying and isolating putative lipopeptide
CC -producing microorganisms. This polynucleotide sequence represents the
CC 88421nt genomic DNA of a ramoplanin producing Actinoplanes sp.
CC
XX

SO Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 other;

Query Match 4.9%; Score 86; DB 24; Length 88421;
Best Match Similarity 43.1%; Pred. No. 2e-05; Mismatches 775; Indels 24; Gaps 3;

Matches 605; Conservative 0; Mismatches 775; Indels 24; Gaps 3;

QY 35 TGTCCCAATCCAGATTCACACCCCTTGTGCTGCTGCGCGGCGACCCCTCATGACG 94
DB 46230 TGGCCCGGCGAGGTGCTGCGACACGCTGTACGCTGCGCGGCGACCCGCTGCAAC 46239
QY 95 GCAGCAACACCCCGGCG 154
DB 46290 ACATCTTCACACCCCGGCG 46349
QY 155 TCGGCGATCTGTGAGACG 214
DB 46350 GGGTCGAGAGGCGGCG 46409
QY 215 CGCGCGCTTCAATCAGATCGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 268
DB 46410 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 46469
QY 269 GCGACATGACGCCCAAGATCTGCGAGGCGGTACACGCTGTACAGGCAATTCGCGCA 328
DB 46470 ACCACATCG 46529
QY 329 TCAGCTGCG 388
DB 46530 ACG 46589
QY 389 GCGGCTTTCACGCTTTCAGAGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 448
DB 46590 ACTATGCGCTTTCGAGCGGTGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46649
QY 449 CGGCGCTCAACG 508
DB 46650 CCGCGCAATCG 46709
QY 509 ACTTGCACG 568
DB 46710 CCGCGCACCG 46769
QY 569 CCATGCGACG 628
DB 46770 TCCG 46829
QY 629 CCACCGCGAGAGATCATGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 688

```

/product- "EvDA"
  complement (1389..1394)
  /tag- b
  complement (1490..2611)
  /tag- c
  /product- "EvDB"
  complement (2618..2622)
  /tag- d
  complement (2622..3860)
  /tag- e
  /product- "EvDC"
  complement (3867..3870)
  /tag- f
  4143..5312
  /tag- g
  /product- "EvDD"
  4134..4138
  /tag- h
  5309..6235
  /tag- i
  /product- "EvDE"
  6232..7275
  /tag- j
  /product- "EvDP"
  6226..6229
  /tag- k
  7272..8327
  /tag- l
  /product- "EvDG"
  8342..9364
  /tag- m
  /product- "EvDH"
  8333..8336
  /tag- n
  complement (9463..10224)
  /tag- o
  /product- "EvDI"
  complement (10232..10235)
  /tag- p
  10424..11176
  /tag- q
  /product- "EvDJ"
  12027..12455
  /tag- r
  /product- "EvDK"
  /partial
  /note- "No start codon"
  complement (12108..13022)
  /tag- s
  /product- "EvDL"
  complement (13027..13030)
  /tag- t
  complement (14410..15363)
  /tag- u
  /product- "EvEA"
  complement (15369..15373)
  /tag- v
  complement (15380..16414)
  /tag- w
  /product- "EvEB"
  complement 16419..17873
  /tag- x
  /product- "EvEC"
  complement (17870..18934)
  /tag- y
  /product- "EvED"
  19374..20906
  /tag- z
  /product- "EvEY"
  21064..22542
  /tag- aa
  /product- "EvYF"
  1056..22542

```



```

FT CDS /tag- ab
FT 22748..24172
FT /tag- ac
FT /product- "EvrG"
FT 22736..22740
FT /tag- ad
FT complement (24177..25223)
FT /tag- ae
FT /product- "EvrH"
FT complement (25230..25233)
FT /tag- af
FT 25550..26626
FT /tag- ag
FT /product- "EvrI"
FT 26685..30479
FT /tag- ah
FT /product- "EvrJ"
FT 26672..26676
FT /tag- ai
FT complement (30557..31876)
FT /tag- aj
FT /product- "EvrK"
FT complement (31885..31888)
FT /tag- ak
FT complement (31941..32882)
FT /tag- al
FT /product- "EvrL"
FT complement (33167..34405)
FT /tag- am
FT /product- "EvrM"
FT complement (34414..34418)
FT /tag- an
FT complement (34449..35210)
FT /tag- ao
FT /product- "EvrN"
FT complement (35219..35221)
FT /tag- ap
FT complement (35294..36238)
FT /tag- aq
FT /product- "EvrO"
FT complement (36235..36963)
FT /tag- ar
FT /product- "EvrP"
FT complement (36998..38026)
FT /tag- as
FT /product- "EvrQ"
FT complement (38072..38566)
FT /tag- at
FT /product- "EvrR"
FT complement (38992..40163)
FT /tag- au
FT /product- "EvrS"
FT complement (40216..40890)
FT /tag- av
FT /product- "EvrT"
FT complement (40899..40902)
FT /tag- aw
FT complement (40887..41576)
FT /tag- ax
FT /product- "EvrU"
FT complement (41679..42707)
FT /tag- ay
FT /product- "EvrV"
FT complement (42714..42717)
FT /tag- az
FT complement (42810..43799)
FT /tag- ba
FT /product- "EvrW"
FT complement (43807..43811)
FT /tag- bb
FT complement (43799..44866)
FT /tag- bc
FT /product- "EvrX"

FT CDS complement (45014..45760)
FT /tag- bd
FT /product- "EvrY"
FT complement (45767..45770)
FT /tag- be
FT complement (45962..46714)
FT /tag- bf
FT /product- "EvrZ"
FT complement (45952..45956)
FT /tag- bg
FT complement (47156..49234)
FT /tag- bh
FT /product- "EvrA"
FT 51627..52715
FT /tag- bi
FT /product- "EvrB"
FT 51629..51622
FT /tag- bj
FT 52889..53557
FT /tag- bk
FT /product- "EvrC"
FT 53554..54207
FT /tag- bl
FT /product- "EvrD"
FT complement (54362..55117)
FT /tag- bm
FT /product- "EvrE"
FT complement (55125..55128)
FT /tag- bn
FT complement (55135..56094)
FT /tag- bo
FT /product- "EvrF"
FT complement (56100..56103)
FT /tag- bp
FT complement (56184..56813)
FT /tag- bq
FT /product- "EvrG"
FT 56961..58709

Query Match 4.8%; Score 84.6; DB 22; Length 109519;
Best Local Similarity 45.1%; Pred. No. 3.2e-05;
Matches 542; Conservative 0; Mismatches 629; Indels 30; Gaps 5;

QY 117 CGCGACCTGGGCGTGGCGGACCGCATCGCCGATCGCGCATCTGTCGAGCGCGC 176
DB 94498 CGTGGACCTTTCACAGGCGCGCGACCTGCGCTTCGTCGACCGGTACGCGCT 94557
QY 177 CGGCGACACCGCGGTGACGTGCGGCTGCGGTGCGCGCGCGCTTCATCGACTCGCA 236
DB 94558 GGAACCGCGCGCGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 94617
QY 237 CACCCACAGACACATACCTGCTCAGGCGTGGCGACATGACGACCGCCAGATCTGCAGGG 296
DB 94618 CGGCTTCGCGGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94677
QY 297 CGTCAACAGGTGTGACGCGCGCAATTGGCGCATCAGCTGCGCGCGCGCGCGCGCA 356
DB 94678 GGTCAACCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94737
QY 357 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 404
DB 94738 CTTGGCGCGCGCGTGAAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94797
QY 405 CGAGCGCTTCGCGCTACCTGACGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 464
DB 94798 CGGCGCGCGCGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94857
QY 465 CTGATGCTGGCGCATTCACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 524
DB 94858 CGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94917
QY 525 CACGACGAGGAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 584

```

Db 94918 CTGACGCGGTACACCGGCGGTACCGGGCCGCGACCGCCCGCCCTGACCGGGTAC 94977
 Oy 585 CATGGGATTCGACCGCGGCTCTACCCCGCCGCGCCGACCAACGAGAT 644
 Db 94978 CTGACCTGCGCGCGCGCGCGCGGTGCGGTGCGCGCGCGCGCGCGAG 95037
 Oy 645 CATGAGGTGTGCGCGCGGTGACCGCGCATGCGCGCATGACGCGCAATGCGCGA 704
 Db 95038 CACCTGCGCGCGGTCTTCCACCGCGACGCGCGCGCGCGCGCGCGCGCGCG 95097
 Oy 705 CGAAGCGGACATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 764
 Db 95098 CGGCG 95157
 Oy 765 CGTCCGCGGTGTGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 818
 Db 95158 CGCGGAGCGGTACTCTTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95217
 Oy 819 CTGCGCGGAGCGGTGCGCGCGGTGATCGAGCGCGCGCGCGCGCGCGCGCGCGCG 878
 Db 95218 CGCGGAGCGGAGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95277
 Oy 879 CGGATATCTGCG 934
 Db 95278 CGCGGAGCG 95337
 Oy 935 CGGAGCG 992
 Db 95338 CGAGCG 95397
 Oy 993 GATGAAAGTGTGCG 1046
 Db 95398 CGAGCG 95457
 Oy 1047 GCAGCG 1106
 Db 95458 GCG 95517
 Oy 1107 GTTCG 1166
 Db 95518 CGAGCG 95577
 Oy 1167 CCGTGTGCG 1226
 Db 95578 CGAGTGTGTGCG 95637
 Oy 1227 GCGGAGCG 1286
 Db 95638 GCGCGCGGTACTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95697
 Oy 1287 G 1287
 Db 95698 G 95698

RESULT 12
 AA87286
 ID AA87286 standard; DNA; 1248 BP.
 XX
 AC AA87286;
 XX
 DT 05-JUN-2000 (first entry)
 XX
 DE S. venezuelae desosamine biosynthetic gene desl, SEQ ID NO:7.
 XX
 KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolemia; crop protection agent; ds.
 XX
 OS Streptomyces venezuelae ATCC15439.
 XX

Key Location/Qualifiers
 CDS 1..1248
 FT /*tag= a
 FT /product= "Desl"
 XX
 PN NC020000620-42.
 XX
 XX
 PD 06-JAN-2000.
 XX
 PF 25-JUN-1999; 99MO-US14398.
 XX
 PR 26-JUN-1998; 98US-0105537.
 XX
 PA (MIND) UNIV MINNESOTA.
 XX
 PI Sherman DH, Liu H, Xue Y, Zhao L.
 XX
 DR WPI: 2000-160679/14.
 XX
 DR P-PSDB: AAY7181.
 XX
 PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
 PT synthesis of methymycin and pikromycin.
 XX
 PS Claim 3; Page 353-354; 438pp; English.
 XX
 CC The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the eryc gene cluster of *Saccharopolyspora erythraea* or
 CC streptomycetes antibioticus. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the
 CC production of biologically active macrolides. The macrolide biosynthetic
 CC proteins are useful for synthesis of methymycin, pikromycin,
 CC neomethymycin and narbomycin. The alternative termination of polyketide
 CC synthetase may be useful to prepare novel antibiotics and
 CC polihydroxyalkanoate (PHA) monomers. The compounds produced by the
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 CC biological applications, to engineer PHA monomer synthases or to prepare
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC active against a variety of organisms, e.g., bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polyketides in plants.
 CC Sequences AA87286-287294 represent desosamine biosynthetic genes from
 CC Streptomyces venezuelae ATCC 15439, which encode proteins
 CC AAY7181-Y7189.
 XX
 SO Sequence 1248 BP; 177 A; 546 C; 377 G; 148 T; 0 other;
 XX

Query Match 4.7%; Score 81.8; DB 21; Length 1248;
 Best Local Similarity 44.2%; Pred. No. 0.00015;
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;
 Oy 32 CCATGTCCCATCCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 91
 Db 215 CCACTGTGACG 274
 Oy 92 ACGGAGCAACACCG 151
 Db 275 AAGTATATATGCG 334
 Oy 152 CCAATGCGGATCTGTGCG 211
 Db 335 TCACCGCGGTCTGCG 394
 Oy 212 TCGGCGCGCGGTTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 271

1	Key	Location/Qualifiers
2	FT	1..1248
3	FT	/*tag- a
4	FT	/product- "Streptomyces venezuelae Desi protein"
5	FT	/transl_except- (pos:1..3, aa:Met)
6	FT	/note- "CDS does not include start codon"
7	FT	/partial
8	FT	
9	FT	
10	FT	
11	FT	
12	FT	
13	FT	
14	FT	
15	FT	
16	FT	
17	FT	
18	FT	
19	FT	
20	FT	
21	FT	
22	FT	
23	FT	
24	FT	
25	FT	
26	FT	
27	FT	
28	FT	
29	FT	
30	FT	
31	FT	
32	FT	
33	FT	
34	FT	
35	FT	
36	FT	
37	FT	
38	FT	
39	FT	
40	FT	
41	FT	
42	FT	
43	FT	
44	FT	
45	FT	
46	FT	
47	FT	
48	FT	
49	FT	
50	FT	
51	FT	
52	FT	
53	FT	
54	FT	
55	FT	
56	FT	
57	FT	
58	FT	
59	FT	
60	FT	
61	FT	
62	FT	
63	FT	
64	FT	
65	FT	
66	FT	
67	FT	
68	FT	
69	FT	
70	FT	
71	FT	
72	FT	
73	FT	
74	FT	
75	FT	
76	FT	
77	FT	
78	FT	
79	FT	
80	FT	
81	FT	
82	FT	
83	FT	
84	FT	
85	FT	
86	FT	
87	FT	
88	FT	
89	FT	
90	FT	
91	FT	
92	FT	
93	FT	
94	FT	
95	FT	
96	FT	
97	FT	
98	FT	
99	FT	
100	FT	
101	FT	
102	FT	
103	FT	
104	FT	
105	FT	
106	FT	
107	FT	
108	FT	
109	FT	
110	FT	
111	FT	
112	FT	
113	FT	
114	FT	
115	FT	
116	FT	
117	FT	
118	FT	
119	FT	
120	FT	
121	FT	
122	FT	
123	FT	
124	FT	
125	FT	
126	FT	
127	FT	
128	FT	
129	FT	
130	FT	
131	FT	
132	FT	
133	FT	
134	FT	
135	FT	
136	FT	
137	FT	
138	FT	
139	FT	
140	FT	
141	FT	
142	FT	
143	FT	
144	FT	
145	FT	
146	FT	
147	FT	
148	FT	

[illegible]

CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
 CC protection agents (e.g., fungicides or insecticides) via expression of
 CC polyketides in plants. The present sequence represents the desomine
 CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC
 CC 15439, as given in figure 32.

XX
 SQ Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;

Query Match 4.7%; Score 81.8; DB 21; Length 13613;
 Best Local Similarity 44.2%; Pred. No. 0.00011;
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

QY 32 CCATGTCCTCCATCCGATTCAGCCCTTGGACCTGCTGCGGGGCGGACCTTCATCG 91
 DB 12669 CCACCTGCAACGCGCACGCGCGGCTGCACTCTGCGCACGCGCGGCTCACCGGCG 12610
 QY 92 ACGGCAAGCAACACCGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 151
 DB 12609 AAGTATCATGCGCGTGTGATGAGCTTGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 12550
 QY 152 CCATGCGGCGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 211
 DB 12549 TCACCGCGGCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12490
 QY 212 TCGCGCGCGGCTTCATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 271
 DB 12489 CCGCGCGGCTTCATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12430
 QY 272 ACATGACCGCCCAATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 331
 DB 12429 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12370
 QY 332 GCGTGGCGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 391
 DB 12369 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12310
 QY 392 GCTCTTACCGTTTGGAGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 451
 DB 12309 AAGTCTTCAGCTTCCA---GCGCACCAAGGCGGCGGCGGCGGCGGCGGCGGCGG 12253
 QY 452 CCGTCAACGCGCGCTGTATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 511
 DB 12252 TCACCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12193
 QY 512 TGCAGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571
 DB 12192 TGC CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12133
 QY 572 TGGCGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631
 DB 12132 TGGGCGCTCACTCTCCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631
 QY 632 CCACGGAAGATCATGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12073
 DB 12072 CCGCGCTACCGCGGAGCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12013
 QY 692 CCCACATGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 751
 DB 12012 ACGGCGCTCAACAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11953
 QY 752 GCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 811
 DB 11952 ACCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 11893
 QY 812 TCGGCGCGCTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 862
 DB 11892 GCGCGGCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 11833
 QY 863 AAGAGCTGCTGCGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 922
 DB 11832 CCGAAGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11773
 QY 923 GCGTCTGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 982

DB 11772 ACATCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11713
 QY 983 GCGCGAGCTGATGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1015
 DB 11712 CCGCGCACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11680

Search completed: May 11, 2003, 01:57:18
 Job time: 1433 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 11, 2003, 12:09:25 ; Search time 85 Seconds

(Without alignments) 8523.078 Million cell updates/sec

Title: US-10-009-782-1
Perfect score: 3299
Sequence: 1 gattccactgacgcgcga.....ccctgactacgagaagcct 1758

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_n2p_model -DEV=xlp
-O=/gen2.1/USPC/spool/US10009782/runat_07052003_122517_23135/app_query.fasta.1.1927
-DB=SPREMBL_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsun62 -TRANS=human40.ccl -LIST=45
-DOCALLIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MTELEN=0 -MATELEN=2000000000
-USER=US10009782.ecgn.1.1.99.grunat_07052003_122517_23135 -NCPV=6 -ICPV=3
-NO.XIPYX -NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL_21:*

- 1: sp_athea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organella:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	2216	67.2	484 2	Q9AGH8 alcaigenes

2	1064	32.3	494	16	Q8XON5	Q8xg5 ralsstonia s
3	942	28.6	526	17	Q9V2D3	Q9v2d3 pyrococcus
4	656.5	19.9	536	16	Q93JK8	Q93jk8 streptomyces
5	407	12.3	581	16	Q9A2U6	Q9a2u6 caulobacter
6	368	11.2	608	12	Q905K9	Q9q5k9 herpesvirus
7	339	10.3	890	12	Q8BZE1	Q8bze1 cercopithec
8	338.5	10.3	529	2	P94809	P94809 microbacter
9	337	10.2	924	12	Q99307	Q99307 epstein-bar
10	332	10.0	3084	12	Q8U211	Q8u211 pseudodora
11	332	9.9	608	12	Q905K9	Q9q5k9 herpesvirus
12	318	9.8	890	12	Q8BZE1	Q8bze1 cercopithec
13	307	9.3	680	2	Q52747	Q52747 ruminoococ
14	306	9.3	586	6	Q95UD1	Q95ud1 sus scrofa
15	299	9.2	706	12	Q41972	Q41972 murid herpe
16	299	9.2	727	12	Q41973	Q41973 murid herpe
17	299	9.2	774	12	Q41971	Q41971 murid herpe
18	296	9.1	1791	4	Q60382	Q60382 homo sapien
19	295	9.1	2752	4	Q9U035	Q9u035 homo sapien
20	295	9.1	1275	4	Q9U036	Q9u036 homo sapien
21	295	9.1	1783	4	Q15038	Q15038 homo sapien
22	293	8.9	676	6	Q95JDO	Q95jdo sus scrofa
23	292.5	8.9	511	6	Q9BKV7	Q9bkv7 leishmania
24	288.5	8.7	1325	5	Q53913	Q53913 streptomyces
25	286.5	8.7	527	2	Q53917	Q53917 nephila mad
26	286.5	8.7	1953	5	Q9B1T7	Q9b1t7 homo sapien
27	285.5	8.6	1404	4	Q9B499	Q9b499 zea mays (m
28	284.5	8.6	1315	10	Q9SPM0	Q9spm0 zea mays (m
29	284	8.6	1404	4	Q9B499	Q9b499 zea mays (m
30	284	8.6	3247	12	Q65553	Q65553 bovine hecp
31	280	8.5	406	2	Q8VPM0	Q8vpm0 halobacteri
32	277.5	8.4	494	17	Q9HNR5	Q9hnr5 halobacteri
33	274.5	8.3	1453	11	Q63079	Q63079 rattus norv
34	272.5	8.3	1414	5	Q26634	Q26634 strongyloce
35	272	8.4	1262	4	Q9U040	Q9u040 homo sapien
36	272	8.4	2295	4	Q9U048	Q9u048 homo sapien
37	271	8.2	787	2	Q8VPM1	Q8vpm1 micrococcus
38	271	8.2	956	4	Q9U039	Q9u039 homo sapien
39	271	8.2	1262	4	Q9U040	Q9u040 homo sapien
40	271	8.2	2752	4	Q9U035	Q9u035 homo sapien
41	270.5	8.2	507	13	Q13028	Q13028 boreogadus
42	270	8.2	1461	4	Q76045	Q76045 homo sapien
43	268.5	8.3	601	5	Q62185	Q62185 caenorhabd
44	268.5	8.1	2296	4	Q9U048	Q9u048 homo sapien
45	267.5	8.1	1445	13	Q93251	Q93251 rana catesb

ALIGNMENTS

RESULT 1

Q9AGH8 PRELIMINARY; PRT; 484 AA.

AC 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE D-aminocyclase (EC 3.5.1.81).

OS Alcaligenes faecalis.

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

OC Alcaligenes.

OX NCBI_Taxid=511;

RN [1]

RP SEQUENCE FROM N.A.

RA Hsu C.S., Lai W.L., Chang W.W., Yang Y.B., Tsai Y.C.;

RT "Gene cloning, nucleotide sequencing of Alcaligenes faecalis DAI D-

RT aminocyclase";

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL, AF332546; AAK15530.1; -

KW Hydrolase.

SO SEQUENCE 484 AA; 52000 MW; 634119A6CB75155E CRC64;

Alignment Scores:

Pred. No.: 4.96e-104 Length: 484

Score: 2216.00 Matches: 416


```

QY 52 CAGCCCTTCAGCTGCTGCTGCGGGGCGGACCCCTCATGACGCGACGACACCCCGGG 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 GlnGlnTyrAspSerValIleArgArgValArgIleValAspArgIleGlnPro 21
QY 112 CCGCGCGCC-----GACCTGGGCGTGGCGGCGGACCGCATCGCCGATCGCGGATG 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22 GluSerThrLeuPheAspValAlaIleAlaAspIleArgIleAlaIleAlaIleAla---Thr 40
QY 166 TCGGACCGCGCGCG-----CACACCGCGGTCGACGTGGCGGCGTGGTGGCGG 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 SerAspSerThrAlaThrPheGlnIleAspIleValAspGlnIleArgValLeuAla 60
QY 217 CCGCGCTTCAGCTGACGCGACACCGACGACGACGACGACGACGACGACGACGACG 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ProGlnThrIleAspValIleThrIleAspAspThrAspValIleArgThrProAspMet 80
QY 277 ACCGCGCAAGATCTGCGAGCGGCGTACACGAGGCGTACGCGGCGATTCGCGGATG 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 LeuProIleValSerGlnGlnIleThrValIleValGlnIleValGlnIleValGlnIleVal 100
QY 337 GCGCGCGTGGCGG---CAGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 SerProIleThrIleLeuArgGlnIleProIleProIleProIleProIleProIlePro 120
QY 394 TCTTACCGTTCGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 AlaPheArgTyrProThrIlePheIleSerTyrValGlnIleValGlnIleValGlnIleVal 140
QY 454 GTCAACCGCGCGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 ValAspValAlaIleValGlnIleValGlnIleValGlnIleValGlnIleValGlnIleVal 160
QY 514 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161 AspArgProIleThrIleProGlnIleValGlnIleValGlnIleValGlnIleValGlnIleVal 180
QY 574 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 AspIleGlnIleValGlnIleValGlnIleValGlnIleValGlnIleValGlnIleValGlnIleVal 200
QY 634 ACCGACGACGATCATGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 ThrGlnIleValIleMetSerIleAlaGlnIleProIleValGlnIleValGlnIleValGlnIleVal 220
QY 694 CACATCGCGCGAGCGAGCGAGCGACATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 221 HisLeuArgThrGlnPheAlaIleIleValGlnIleValGlnIleValGlnIleValGlnIleVal 240
QY 754 CCGGAGCTGACGTGCGCGCGTGTATCTCGACACCAAGGTTCATGGCGCGCGCGCGCGCG 813
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 ArgHisIleAlaArgValProIleValIleSerHisIleValGlnIleValGlnIleValGlnIleVal 260
QY 814 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 GlnArgIleValGlnIleValGlnIleValGlnIleValGlnIleValGlnIleValGlnIleVal 280
QY 874 CTGAGCGCGGTATCCCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 CysAspArgTyrProIleThrIleSerSerThrIleAspLeuValGlnIleValThrAsp 300
QY 934 GCGCGAGCGACCATCATCTGCTGCAAGCGCTTCGCCGCAAGTGAAGCGCGCGCGCGCGCG 993
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GluPheAspIleValIleThrIleThrSerAspProHisProGlnIleValGlnIleValGlnIleVal 320
QY 994 GATGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 LysAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAla 340
QY 1054 GCGCGCGCGCATCTTCATGATGAGCGAGACCGCGACGCGCGCGCGCGCGCGCGCGCGCG 1113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 AlaGlnIleValIleTyrHisAspMetAlaIleAlaIleValGlnIleValGlnIleValGlnIleVal 360

```

```

QY 1114 CCGACCATGATGCGCTCGCGAGCGCGCTGCGCGACGACGACGCGCGCGCGCGCGCG 1173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 AlaThrValIleGlnIleSerAspGlnIleProIleProIleProIleProIleProIlePro 380
QY 1174 GCGACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 GlnAlaPheProHisValLeuAlaIleIleTyrSerArgIleValGlnIleValGlnIleValGlnIleVal 400
QY 1234 ACCGCGGTATGGAAGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 AlaAlaValAlaArgIleMetThrIleGlnIleSerAlaIleValGlnIleValGlnIleValGlnIleVal 420
QY 1294 CAGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 LeuValArgGlnIleTyrThrAlaAspLeuValLeuPheAspProAspThrValArgAsp 440
QY 1354 ACCGCGACCTTCGGAACCGCTACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 LeuAlaThrPheThrAspProIleValGlnIleValGlnIleValGlnIleValGlnIleValGlnIleVal 460
QY 1414 GCGCGCGCGCTGCGACAGCGAGCGGTTCACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 GlnAlaIleValIleTyrArgAspIleValGlnIleProIleValGlnIleValGlnIleValGlnIleVal 479
QY 1474 CCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 -----ArgArgIlePro-----ArgAlaHisAla 487
QY 1534 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1551
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 GlnProIleProGlnIleu 493

RESULT 3
Q9V2D3 PRELIMINARY; PRT; 526 AA.
ID Q9V2D3 AC Q9V2D3
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE D-aminoacylase (Aspartate, glutamate etc).
NM DAD OR PAB0090.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RA Helwig R.;
RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome
   structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ248283; CABA9065.1;
DR InterPro: IPR002832; Adenine_deam.
KW pfam; PF01979; Adenine_deam; 1.
KW Complete proteome.
SQ SEQUENCE 526 AA; 58790 MW; 705B1B7B281915A0 CRC64;

Alignment Scores:
Pred. No.: 6.49e-40 Length: 526
Score: 942.00 Matches: 213
Percent Similarity: 55.56% Conservative: 82
Best Local Similarity: 40.11% Mismatches: 172
Query Match: 28.55% Indels: 64
DB: 17 Gaps: 9

US-10-009-782-1 (1-1758) x Q9V2D3 (1-526)
QY 58 TTCGACCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 TyrAspIleValIleTyrAsnGlnIleValIleValAspGlnIleValAspProIleThrPheArg 23
QY 118 GCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 177

```


DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE LF3.
 OS cercopithecine herpesvirus 15.
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Gammaherpesvirinae; Lymphocryptovirus.
 NCBI_TaxID=104228;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RX MEDLINE-97048062; PubMed-8892903;
 RA Franken M., Devergne O., Rosenzweig M., Annis B., Kieff E., Wang F.;
 RT "Comparative analysis identifies conserved tumor necrosis factor
 receptor-associated factor 3 binding sites in the human and simian
 Epstein-Barr virus oncogene LMP1.";
 RL J. Virol. 70:7819-7826(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RX MEDLINE-99412410; PubMed-10482645;
 RA Rivallier P., Quink C., Wang F.;
 RT "Strong selective pressure for evolution of an Epstein-Barr virus
 LMP2B homologue in the rhesus lymphocryptovirus.";
 RL J. Virol. 73:8867-8872(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RX MEDLINE-20304984; PubMed-10846073;
 RA Jiang H., Cho Y.-G., Wang F.;
 RT "Structural, functional, and genetic comparisons of Epstein-Barr virus
 nuclear antigen 3A, 3B, and 3C homologues encoded by the rhesus
 lymphocryptovirus.";
 RL J. Virol. 74:5921-5932(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RX MEDLINE-20440633; PubMed-10970361;
 RA Rao P., Jiang H., Wang F.;
 RT "Cloning of the rhesus lymphocryptovirus viral capsid antigen and
 Epstein-Barr virus-encoded small RNA homologues and use in diagnosis
 of acute and persistent infections.";
 RL J. Clin. Microbiol. 38:3219-3225(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RX MEDLINE-21602573; PubMed-11739708;
 RA Rivallier P., Jiang H., Cho Y.-G., Quink C., Wang F.;
 RT "Complete Nucleotide Sequence of the Rhesus Lymphocryptovirus: Genetic
 Validation for an Epstein-Barr Virus Animal Model.";
 RL J. Virol. 76:421-426(2002).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RX Moghaddam A., Koch J., Annis B., Wang F.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RX Moghaddam A., Annis B., Wang F.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RX Rivallier P., Quink C., Wang F.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RX Jiang H., Wang F.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [10]
 RP SEQUENCE FROM N.A.

RC STRAIN-LCL8664;
 RA Rao P.V., Jiang H., Wang F.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [11]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LCL8664;
 RA Rivallier P., Jiang H., Cho Y.-G., Quink C., Wang F.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A1037858; AAK95470.1;
 SQ SEQUENCE 890 AA; 95097 MW; 149077C89D3B5574 CRC64;

Alignment Scores:

Pred. No.:	1,53e-09	Length:	890
Score:	339.00	Matches:	199
Percent Similarity:	35.69%	Conservative:	43
Best Local Similarity:	29.35%	Mismatches:	259
Query Match:	10.28%	Indels:	178
DB:	12	Gaps:	40

US-10-009-782-1 (1-1758) x Q8UZE1 (1-890)

QY	51	CGAGCCCTTGACCTGCTGCTGC	---GGGCGGACCCCTGATGAGCGACACACCC	107
DB	28	ProAlaProGlyProSerSerArgThrGlyGlnArgThrProGlyHisPro	147	
QY	108	GGGGCGGCGGCGACCTGGG	---CGTGGCGGCGACCGCATCGCCCATCGGCATCT	164
DB	48	ProProGlyArgGlySerGlyProArgGlyThrArgProAlaProSerArgThr	67	
QY	165	GTCGGACGCGCGCGGACACCGGAT	-----CGAGCTGTGGGCGCTGAT	209
DB	68	GlyGlnArgThrArgGlnArgProGlyHisProProGlyArgGlySerGlyProArg	87	
QY	210	GGTCCGCC	-----CCGGCTTATGACTCGACACCCACGACGACACTACTCTCAGGC	265
DB	88	GlyThrArgProProAlaProSerSerArgThrGlyGlnArgThrArgGlnArgPro	107	
QY	266	GTCGCGACATGACGCCAAGATCTGCGAGCGGCTGACACGAGTGCAGGCAATTGCG	325	
DB	108	HisPro-----ProProGlyArgGlySerGlyProArgGlyHisProProAla	124	
QY	326	GCATCAAGCTGGCGCGCTGGCGACG	-----CCAACCGCGCGCGCTGGACG	376
DB	125	ProSerSerArgThrGlyGlnArgThrArgGlnArgProGlyHisProProPro	142	
QY	377	TGCTGGAGCAAGCGCGCTTACCGCTTCAGCGCTTCGCGACTGACGACGCTGC	436	
DB	143	-----GlnArgGlySerGlyProArgGlyThrArgProProAlaProSerArgThr	160	
QY	437	GGG-----CCAGCGCGCGCGCGCTGACACCGCGCTGTATGGTGGCGCATTCAGCGCTGC	490	
DB	161	GlyGlnArgThrArgGlnArgProGlyHisProPro-----ProGlyArgGlySer	177	
QY	491	GGCGCGCGGCTATCGCGGACTGACGCGG	-----ProGlyArgGlySer	520
DB	178	GlyProArgGlyThrArgProProAlaProSerSerArgThrGlyGlnArgThrArgGln	197	
QY	521	-----CCGCGACGAGAGGAATGAGCGGCGGACGCGGCGGAGAG	568	
DB	198	ArgProGlyHisProProProGlyArgGlySerGlyProArgGlyHisProPro-----	215	
QY	569	CGATGCGCAGCGCGCGCATCGGATTCGACCGCGCGCTTCACCGCGC	616	
DB	216	-----ProAlaProSerSerArgThrGlyGlnArg	225	
QY	617	-----CGCGCGCC	631	
DB	226	ThrArgGlnArgProGlyHisProProProGlyArgGlySerGlyProArgGlyThrArg	245	
QY	632	CCACGAGAGATCATGCGAGTGTCGCGCGCGTGGAGCGCATGGCGCATTCAGCA	691	
DB	246	ProProAlaProSerSerArgThrGlyGlnArgThrArgGlnArgProGlyHisProPro	265	

```

QY 692 CCCACATGCGGAGAGGAGGAGCATGTCGCGCGCTGGAGAAACCTTCCGATCG 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 Pro-----GluArgGlySerGlyProArgGlyThrArgProProAlaPro 280
QY 752 GCCCGAGACTGAGAGTGGCGGTGTGATCTGCACG-----ACAAGGATGAGGCG 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 Ser-----SerArgThrGlyGlnArgThrArgGlnArgPro 292
QY 803 AGCCCAATTTGGCGCGCTGCGCGAGAGAGTGCCTGATCGAGGCGCCGATCG---CGC 859
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 GlyHisProProProGluArgGlySerGlyProArgGlyThrArgProProAlaPro 312
QY 860 GCCAGGAGCTGCTCGGTGAGCGGT---ATCCCTAGCTGGCGCGCGCCCAATCGCATCG 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 SerArgThrGlyGlnArgThrArgGlnArgProGlyHisProProProGluArgGlySer 332
QY 917 AGGACCGCGGTGCTGCTGCGCGAGCGCATCATCA----- 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 333 GlyProArgGlyThrArgProProAlaProSerSerArgThrGlyGlnArgThrArgGln 352
QY 953 ---CCTGCT----- 970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 ArgProGlyHisProProProGluArgGlySerGlyProArgGlyThrArgProProAla 372
QY 971 CGGAACGTAGCGGCGGAGCGAGCATGATGAGTGAAGTGCAGCGCGGCAAAATCGCAATCG 1030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 ProSerSerArgThrGlyGlnArgThrArgGlnArgProGlyHisProProProGluArg 392
QY 1031 ACGTGTGCTCGGAGCTGAGCGCG---CCGGGCGCATCTACTCATGATGGA---CGAAC 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 GlySerGlyProArgGlyThrArgProProAlaProSerSerArgThrGlyGlnArgThr 412
QY 1085 CCGAGCTGACGAGCGCATCTGCGCGCTGCGCGAGCATGATGCGCTCGAGCGCGCTGCGCG 1144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 ArgArgGln-----ArgProGly---HisProProProGluArgGlySerGlyProArgGly 429
QY 1145 AGGAGAGCGCGCGATCGCGGCT-----GTGGGCGCA 1177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 YThrArgProProAlaProSerSerArgThrGlyGlnArgThrArgGlnArgProGlyHis 449
QY 1178 CCGTCCG-----CGGGGTGCTGGGCGCATATGCGCGAGCTGGCGCGCTGCGCGTGG 1231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 449 SProProProGluArgGlySerGlyProArgGlyThrArgPro---ProAlaProSerSe 468
QY 1232 AGACGCGCGTATGAGATGACCGG-----CCTGACCGC----- 1265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 ArgThrGlyGlnArgThrArgGlnArgProGlyHisProProProGluArgGlySerGly 488
QY 1266 ---CGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 YProArgGlyThrArgProProAlaProSerSerArgThrGlyGlnArgThrArgGlnArg 508
QY 1319 ACCTGTGTGTGTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 508 GProGlyHisProProProGluArgGlySerGlyProArgGlyThrArgProProAlaPro 528
QY 1379 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 SerSerArgThrGlyGlnArgThrArgGlnArgProGlyHisProProProGluArg 547
QY 1421 CGGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 GglySerGlyProArgGlyThrArgProProAla---ProSerSerArgThrGlyGly 565
QY 1481 CGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 565 nArg-----ThrArgGlnArgProGlyHisProProPro 576
QY 1541 T-----CCCAACCTGAGAGCAAAACCGCTACATGCGCGCTCCCTCCGCGTC 1585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 GglySerGlySerGlyProArgGlyThrArgProPro---AlaProSerSerArg--- 593
QY 1586 GCATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1645

```

```

Db 594 -----ThrGlyGlnArgThrArgGlnArgProGlyHisProProProGluArg 609
QY 1646 AGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 609 gglySerGlyProArgGly-----ThrArgProProAlaProSerSerArgThrGlyGly 627
QY 1691 CGCGCATCGC-----GGCTCGACCGCTGCGCAAGCGCGAGCTGGCGCA 1733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 627 nArgThrArgGlnArgProGlyHisProProProGluArgGlySerGlyPro 644
RESULT 8
P94909 PRELIMINARY: PRT: 529 AA.
AC P94909;
DT 01-MAY-1997 (TREMBLE, 03, Created)
DR 01-MAY-1997 (TREMBLE, 03, Last sequence update)
DE 01-NOV-1998 (TREMBLE, 08, Last annotation update)
DE Hypothetical 58.7 kDa protein (Fragment).
OS Microbacterium ammoniaphilum.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Microbacteriaceae; Microbacterium.
OX NCBI_TaxID=42460;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15354;
RX MEDLINE=96257250; PubMed=8654988;
RA Striebel H.M., Seebler S., Jarsch M., Kessler C.;
RT Cloning and characterization of the Mami restriction-modification
RT system from Microbacterium ammoniaphilum in Escherichia coli.;
RL Gene 172:41-46(1996).
DR EMBL, X79027; CAA55649.1;
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 529 AA; 58716 MW; 5DC1216237DE772D CRC64;
Alignment Scores:
Pred. No.: 1,52e-09 Length: 529
Score: 338.50 Matches: 181
Percent Similarity: 35.59% Conservative: 45
Best Local Similarity: 28.50% Mismatches: 198
Query Match: 10.26% Indels: 211
Gaps: 30
US-10-009-782-1 (1-1758) x P94909 (1-529)
QY 32 CCATGTCCCATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCG 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ProAlaProSerAlaThrArgSerProSerTyrGlySerArgGlyYlleProThr 20
QY 77 -----GGCGACCTCATCG 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 SerProAlaThrThrSerSerThrGlyArgThrAlaProAlaAlaSerAlaProAlaSer 40
QY 92 AGCGGAGCAAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 ThrAlaGluProHisAlaGlyProAla-LysProGlyArgAlaGluProAlaGlyThrGly 60
QY 135 -----CGCGGACCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 YThrIleGluProArgGlySerHisIleTyrGlnHisArgThrArgProAlaGlnArgGlnAla 80
QY 182 ACACCGGCGTCACTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 HisProLeuGlnArgPro-----HisGlnGluProGlyAla 92
QY 242 ACAGAGCACTACTGCTGTCAGCGCTGCGGACATGACGCCCAAGATCTCGAGGCGCTCA 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 nArgGlyGlyLeuProValArgAlaAspArgAlaGlu----- 104
QY 302 CCAGCGTGTGTCAGGCGCAATTCGCGCATGACGCTGCGCGCGCGCGCGCGCGCGCGCG 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 105 -----ArgGlyArgArgGluProAla 111

```

OY	1323	-GCTGGTGTTCGACCCGCC-	-----CACGGTGCGCATCCGCCCACTTCGAAACAC	1377
Db	392	gargtglyproatrgnglmglyglalysllygely	-----glnglyleuaarghisal	410
OY	1373	CACAGAGGCGCCCCCGGCATCCATTCCGTTACGTACGACAAGCGCGCGCGGTCTGGCAG	14322	
Db	410	aqlayrgharghargharglyvalproserglnleualavalLeuarGlnGlyValCySgl	430	
OY	1433	AGCAGCGCTTACC-----GGCAGCATGCCGGCGGGTGTCTGCACGACGAGCGC	1483	
Db	430	nAlaserla-Threulaleutrimptetmrsergglllylryleuauyrgluleuulgag	450	
OY	1484	CCTGAGCCCGCGCCACGCCCTTACAATCGCGCGCTGAACGGGGCGCGTGC GCCCTCC	1543	
Db	450	lyleuLeucysarapsarservalgniser	-----	459
OY	1544	CACCCCTGAGCAGCAACCGCGAATCGCCCTCCCTCC- -GCTCGCAATAAGGCCCCAC	1600	
Db	460	--HisleuaspyalthrThrprrgalaaapalaasarlilesearArgTrh-----	476	
OY	1601	CCGATATCGTGGCGAAGGAMGTATGGCGCGCGCTCGCGCGCGGACCGAGAAGCC--	1656	
Db	477	-----TriphrValmethasnThrArgleuAyglnalaglnlysSerleua	492	
OY	1657	-----CGAATAATGACCTCGCAACACTGTGCGAGGCC	1689	
Db	492	laalaahargprothrThrserserislerarsgrs	505	
RESULT	9			
ID	099307	PRELIMINARY;	PRT; 924 AA.	
AC	099307;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DE	01-FEB-1997 (TREMBLrel. 02, Last annotation update).			
OS	IF3 protein.			
OS	Epsstein-Barr virus (strain Raj1) (Human herpesvirus 4).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Gammaherpesvirinae; Lymphocryptovirus.			
RN	NCBI_TaxID=10378;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Parker B.D., Bankier A., Satchwell S., Barrell B., Farrell P.J.;			
RL	Virology 0:0-(1990)			
DR	EMBL: M35547; AAA45896.1;			
SO	SEQUENCE 924 AA; 94304 MW; 3C7C86297F0E0F4 CRC64;			
Alignment Scores:				
Pred. No.:	1,94e-09	Length:	924	
Score:	337.00	Matches:	210	
Percent Similarity:	34.58%	Conservative:	30	
Best Local Similarity:	30.26%	Mismatches:	287	
Query Match:	10.22%	Indels:	168	
DB:	12	Gaps:	31	
US-10-009-782-1 (1-1758) x 099307 (1-924)				
OY	45	CGATCCGACCCCTTCGACT---GCTGCTGCGGGGCGACCTCATCGACGCACAA	101	
Db	6	ArgglyProCysleuAlaProGlySerIylleucllyalAhSPRoHslPrOAlArGrSer	25	
OY	102	-----CACCCGCGGGGCGCGCGCGCGACCTGGGGCGT	131	
Db	26	glyAlaAlaAsPProAlaAsPProVaIlGlyHisPProAlaAlaPProAlaAlaProGly---	44	
OY	132	CGCGGGGAGCGCATCCCGCGCATCGCGCATCTGTCTGGAGCGCGCGC-----	179	
Db	45	----ProGlupProAlrghThrArgleuGlnProAlaThrProArgrArSerelYalAlaAlaAsP	63	
OY	180	-----GCACACCGCGGr-----CGACGTCTCGGGCTGGGTGGCGCC	218	
Db	64	ProAlaAsPProVaIlGlyHisPProAlaAlaPProAlaAlaProGly---	44	


```

OY 219 CGGCTTCATGAGCTCGACACCCAGACGAACTACCTGCTCAGGGGTCGACATGAC 278
    |||||  ::  |||  |||||  |||  ::|||
Db 84 ArgLeu---GlnProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAsp 102
OY 279 GCCCAAGATGTCGAGGGCTCACCCAGCTGGTCACGGGCAATTGCGCATACCTGCGC 338
    |||  |||  |||  |||  |||  |||  |||  |||
Db 103 GlnHsProAlaAlaProAlaProGlyProGlyProAlaGlySerGlyAlaAlaAsp 122
OY 339 GCC-----GCTGGCGGACGGCAACCCCGCGCGCC----- 368
    ::|||  |||  |||  |||  |||  |||  |||  |||
Db 123 ThrProAlaGlySerGlyAlaAlaAspProAlaAspProValGlyHsProAlaAlaPro 142
OY 369 -----CTGAGACCTGTCGAGAGCGGCTC-----TTACGGTTTCAGCGCTT 413
    |||||  |||  |||||  |||  |||||  |||  |||||  |||
Db 143 ArgAlaProGlyProGlyProAlaGlyHsProAlaThrProAlaGlySerGly 162
OY 414 CGCGGACCTACCTGGAGCGCTGGCGGCGACCGCGCGCTGACACCGCTGTATAGT 473
    |||  |||  |||||  |||||  |||||  |||||  |||||  |||
Db 163 AlaAlaAspProAlaAspProValGlyHs-ProAlaAlaProAlaGlyProGlyProG 182
OY 474 GGGGCTTCAGCGCTCGCGCGCGCTGATCGCGGACCTGACCGCGCGCGCGCGACGA 533
    |||  |||  |||  |||  |||  |||  |||  |||
Db 182 LnpProAlaGlyHsProAlaThrProAlaGlySerGlyAlaAlaAspPro---- 200
OY 534 GGAATCGCGCGCTTCACCTGGCGGCGCGGAGAGCGCATGCGCGCGCGCGCAT 593
    ::  |||  |||  |||  |||  |||  |||  |||
Db 201 -----AlaAspProValGly-----HisProAlaAlaProAlaGlyAla 213
OY 594 TTCGACCGCGCGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641
    |||  |||  |||||  |||||  |||||  |||||  |||||  |||
Db 213 roGlyProGlyProAlaGlyHsProAlaThrPro-AlaGlySerGlyAlaAla 232
OY 642 -----GATCATGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 692
    |||  |||  |||  |||||  |||||  |||||  |||||  |||
Db 233 AspProAlaAspProValGlyHsProAlaAlaProAlaGlyAla-----ProGlyProGly 250
OY 693 CCACATGCGCGCGAGGCGCGACATCGTGGCGCGCGCGCGCGCGCGCGCGCAT 752
    |||  |||  |||  |||  |||  |||  |||  |||
Db 251 ProAlaGlyHsProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAsp 270
OY 753 CGGAGAGCTGAGCGCGCGGTGATCTCGGACCAAGGTCATGCGCGA----- 803
    |||  |||  |||  |||||  |||||  |||||  |||||  |||
Db 271 Pro-----ValGlyHsProAlaAlaProAlaGlyAlaProGlyProGlyProAlaG 286
OY 804 GCCCAATTCGCGCGCTCGCG-----CGAGAGCGTGGCGGTGATGAGGCG----- 848
    ::|||  |||  |||||  |||||  |||||  |||||  |||||  |||
Db 287 ThrArgLeuGlnProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAspProVal 306
OY 849 CGCCATGCGCGCGCGAGGCTCGCTGGA-----CGCGTATCCCTACGTCGCGG 899
    |||  |||  |||  |||  |||  |||  |||  |||
Db 307 GlnHsProAlaAlaProAlaGlyProGlyProGlyProAlaGlyHsProAlaGlyHsPro 326
OY 900 CTCGACCTGCTCAAGCAGGACCGCG-----TGTGCTGTCGCGCGGACGAC 944
    |||  |||  |||  |||  |||  |||  |||  |||
Db 326 LnpProAlaGlySerGlyAlaAlaAspProAlaAspProValGlyHsProAlaAlaPro 346
OY 945 CATATCACCGTGGTCAGAGCGCTTCGCGGACGAGCGCGCGCGCGCGCGCGCAT 1004
    |||  |||  |||  |||  |||  |||  |||  |||
Db 346 roAlaGlyAlaProGlyProGlyProAlaGlyHsProAlaThr-----ProAlaGly 364
OY 1005 GCGCGAGCGCGCAATCAAGTACGAGCGTGGCGCGCGCGCGCGCGCGCGCGCG 1058
    |||  |||||  |||||  |||||  |||||  |||||  |||||  |||
Db 364 rgSerGlyAlaAlaAspProAlaAspPro-ValGlyHsProAlaAlaProAlaGlyAlaPro 383
OY 1059 -----CGCATCTCTTCATGATGAGACGACCGCGCGCGCGCGCGCGCGCG 1106
    |||  |||  |||  |||  |||  |||  |||  |||
Db 384 GlyProGlyProAlaGlyHsProAlaThrProAlaGlySer-----Gly 400
OY 1107 GTTCGCGCGCGCATGATGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1166
    |||  |||  |||  |||  |||  |||  |||  |||
Db 401 AlaAlaAspProAlaAspProValGlyHsProAlaAlaProAlaGlyAlaProGlyProGly 420
OY 1167 CCGTGGGCGACCTT----- 1181

```

```

Db 421 ProAlaGlyHsProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAsp 440
OY 1182 CCGCGCGGTGCTGGGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1241
    |||  |||  |||  |||  |||  |||  |||  |||
Db 441 ProValGlyHsProAlaAlaProAlaGlyAlaProGlyProGlyProGlyProGly 455
OY 1242 ATGGAAGATGACCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1301
    |||  |||  |||  |||  |||  |||  |||  |||
Db 456 -----ArgThrArgLeuGlnProAlaThrProAlaThrProAlaGlySerGly 468
OY 1302 GCGCGGCTACTCGCGGACCTGCTGTGT-TGACCGCGCGCGCGCGCGCGCGCG 1360
    |||  |||  |||  |||  |||  |||  |||  |||
Db 469 AlaAlaAspProAlaAspProValGlyHsProAlaAlaProAlaGlyAlaProGlyProGly 488
OY 1361 CCTCGAAGACCCCA----- 1375
    |||  |||  |||  |||  |||  |||  |||  |||
Db 489 ProAlaGlyHsProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAsp 508
OY 1376 -----CGAGCGCGCGCGCGCGCATTCCTCGGTACGTCAGCGCGCGCGG 1423
    |||||  |||||  |||  |||  |||  |||  |||  |||
Db 509 ProValGlyHsProAlaAlaProAlaGlyAlaProGlyProGlyProAlaGlyHsPro 528
OY 1424 TCTGGCAAGACGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1465
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||
Db 529 ProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAspProValGlyHsProAla 548
OY 1466 TGTCTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1519
    |||  |||  |||||  |||||  |||||  |||||  |||||  |||
Db 549 AlaProAlaGlyAlaProGlyProGlyProAlaGlyHsProAlaThrPro-AlaGly 568
OY 1520 ACGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1564
    |||  |||  |||  |||  |||  |||  |||  |||
Db 568 gSerGlyAlaAlaAspProAlaAspProValGlyHsProAlaAlaProAlaGlyAlaPro 587
OY 1565 ACATGGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||
Db 588 -----GlyProGlyProAlaGlyHsProAlaThrProAlaGlySerGlyAlaAla 606
OY 1625 T-----GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1664
    |||  |||  |||  |||  |||  |||  |||  |||
Db 606 AspProAlaAspProValGlyHsProAlaAlaProAlaGlyAlaProGlyProGlyProAla 626
OY 1665 -----GACCTGCAAGACCTGTCGAGCGCGCGCGCGCGCGCGCGCGCG 1714
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||
Db 626 gThrArgLeuGlnProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAspProVal 646
OY 1715 CCAAGCGCGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1742
    |||  |||  |||  |||  |||  |||  |||  |||
Db 646 LglnHsProAlaAlaProAlaGlyAlaPro 655

```

RESULT 10

08U211 PRELIMINARY; PRT: 3084 AA.

AC 08U211:

DT 01-VAR-2002 (TREMBLrel. 20, Created)

DT 01-VAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-VUN-2002 (TREMBLrel. 21, Last annotation update)

DE UL36 protein.

OS Pseudorabies virus (strain Kaplan) (PRV).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicellovirinae.

OX NCBI_TaxID=33703;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KAPLAN.

RA Klupp B.G., Fuchs W., Granzow H., Nixdorf R., Mettenleiter T.C.;

RT "The Pseudorabies virus UL36 tegument protein physically interacts

RT with the UL37 protein."

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ422133; CADI9511.1; -

DR InterPro; IPR005210; Herpes_UL36.

DR Pfam; PF03586; Herpes_UL36; 1.

SQ SEQUENCE 3084 AA; 324401 MW; ECAD9E1E3DC22D1A CRC64;
 Alignment Scores:
 Score: 4.53e-09 Length: 3084
 Percent Similarity: 331.00 Matches: 187
 Best Local Similarity: 32.72% Conservative: 45
 Query Match: 26.38% Mismatches: 269
 DB: 10.03% Indels: 209
 Gaps: 29
 US-10-009-782-1 (1-1758) x Q8U211 (1-3084)
 14 TCAGGAGAGAGATTCATGTCCTCCAAATCCGAT-----CCAGCCCTTGACG 64
 2253 SerLeuLysAlaArgIleGluProProProProValIleAsnProProTyProAlaThr 2272
 65 TGCTGCTCCGGGGGACCCCTCATGACGGGACGACACCCCGGGGGGGGGCCGAC 124
 2273 -----AlaProAlaProGluThrAlaProProGlu----- 2282
 125 TGAGGCTGGCGGCGACCCATGCGGATCGGCGATGTCGAGC----- 172
 2283 -----AlaProGluAlaGlnProProAlaAlaAlaLysProThrProGluProGln 2299
 173 -----CCGCGCGGACACCCGGGTGAGCTGCGGGCTGTGTCGCGCGCGCT 223
 2300 GlyProProProProGluProGluProSerAlaGlnAlaProProAlaGlnLysProPro 2319
 224 TCATGACTCGACACCCGACGACGACACTACCTGCTCAGGGCTCGGACGACGCGCA 283
 2320 AlaGlnProAlaThrAlaAlaAlaAlaThrAlaProLysAlaThrProGluThrGlnPro 2339
 284 AGATCTCGACGAGGCGTACACGAGTGTACAGGGCAATGCGGCAATGCGGCGCGCG 343
 2340 ProThrArgAla----- 2343
 344 TGCGGACGCGACCCGCGCGCGCGCGCGCGCTGACGAGGAGGCGGCTTACC----- 400
 2344 ---GlnThrGlnThrAlaProProProProSerAlaAlaThrAlaAlaAlaGlnValPro 2362
 401 -----GTTTCAGGCGCTTCCGCGACTACCTGAGCGGCTGGCGGCGCGCGCGCG 454
 2363 ProGlnProProSerSerGlnProAlaAlaLysProArgGlyAlaProProAlaPro 2382
 455 TCAGCGCGCGCTGTATGTGGGCAATCAAGCTGCGCGCGCGCGCGCGCGCGCGCG 514
 2383 AlaProProProPro-----SerAlaGlnThrThrLeuProArgProAlaAlaProPro 2400
 515 AGCGGCGCGCGACGAGGAAATCGCGCGCGCGCGCGCGCGCGCGCGAGGAGCAATGG 574
 2401 AlaProProProProSerAlaGlnThr-----ThrLeuProArg-----ProAla 2415
 575 CCAGGCGCGCGCATTCGCGATTCGACCGCGCGCTTACCGCGCGCGCG-----CCCGCGCA 631
 2416 ProProProProSerAla-----ProAlaAlaThrProThrProProAlaProGlyPro 2433
 632 CCAGGAGAGATCATGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 691
 2434 AlaProSerAlaLysLysSerAspLysAspArgIleValGlnProLysAlaGlyAlaPro 2453
 692 CCCACATGGCGACGAGGCGACATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
 2454 -ProAspValArgAspAlaLysPheGlyGlyLysValAlaGlyAlaLysSerProGlyArg 2473
 752 GCGG-----CGAGCTGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 793
 2473 GProLeuAspAspArgLysArgLysLysAlaGlyAla-----AspLysCysGlnHisThr 2491
 794 TCATGGCGCGCGCAATTCG----- 815
 2491 rValGlyProAspArgLysHisArgLysThrArgSerProProGlyThrSerAspSerGluProAl 2511
 815 ----- 815

Db 2511 aSerAlaSerSerSerAlaAlaSerThrProSerArgGluAlaAlaSerSerSerSe 2531
 816 -----CCGCTCGCGCGAGACGCTGCGCG 838
 2531 rProAlaGlyThrProSerAspLysAlaArgGlyGlyProProSerAlaAlaGlyAlaAl 2551
 839 TCATCGAGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 898
 2551 AlaAlaAlaProProGlyAlaProGlySerGlyGlyGlnArgLysArgLysPheGlyPhe 2571
 899 GCTCCACCATGCTCAAGACGAGGAGCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 958
 2571 AlaAlaAlaSerSerArgLysArgLysPheGlnProAlaGlnHis--LeuProAlaAlaProVal 2590
 959 GCAAGCGCTTCCCGCACTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1018
 2591 AlaGlnProLeuProSerAlaValLeuArgThrLysAlaArgThrSerValGlyLeuArg 2610
 1019 AATCCAGTACGACGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTA 1069
 2611 AspHisSer-ArgArgLysSerMetAlaLeuValProArgArgProValSerAlaProAr 2630
 1070 TCATGATGACGACGACCGGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1129
 2630 GlnThrArgLysLysLeuProProAlaAlaGlnProGluGlyArgProProGlyLeuG1 2650
 1130 CCGAGCGCGCT-----GCGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 1183
 2650 UPProArgProProGlnAlaGlnAlaGlnAlaProAlaThrAlaProThrGlyPro--Pr 2669
 1184 CCGGGTGTGGGGGCGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1240
 2669 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 2689
 1241 TATGGAAGATGACCGGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 1300
 2689 Y-----ProValArgArgAlaAlaGlyPro-----ArgProProAl 2701
 1301 AGCGCGGTA-----CTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1333
 2701 aGlnHisLeuArgLysHisIleLeuProProGluProGlnProValGlnLeuGlnLysValAlaVal 2721
 1334 ACCCGG-CCAGGTGCGCG-----ATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1389
 2721 LProLeuProAlaSerProGluThrProAlaProAlaGlnThrGlnProProArgSerPr 2741
 1390 GGCATCATTCG----- 1402
 2741 OleuAlaAlaProThrSerLeuAlaAlaProSerGluIleGluArgProAlaAlaSerAl 2761
 1403 -----TGACGTCAACG 1440
 2761 AlaAlaAlaAlaAlaAlaAlaThrThrThrThrThrThrThrThrThrThrThrThrThr 2781
 1441 TTCACCG 1494
 2781 rAlaProAlaAlaProProProAlaProProProProSerAlaProAlaAlaProProProAl 2801
 1495 CGCCAGCGCTTCAATCG 1531
 2801 aProProProSerAlaArgProAlaProThrProGluThrGlnHisGlyGluProProArgLeuHe 2821
 1532 -----GCCG 1551
 2821 rPheThrProProProLeuProProGluProProSerGlnGlnArgProProGluAlaProTr 2841
 1552 GACGCGAAGCGGTACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1593
 2841 P-----ThrTrProGluProArgAspSerAlaGlyProProAlaLeuPheThr 2857
 1594 GCCCGACCGGATATGTCGGGCGAGGAGTGAATGATGAGGCGCGCGCGCGCGCGCGCGAG 1653
 111 111 111

```

Db 2857 rProProPro-----SerAsProProValGlnPh 2867
QY 1654 GCCGGGAAATGACCTTCGACAGACTGTGCGAGCGCCAGCGCATGCGGCTTCGACCTG 1713
Db 2867 eProGly-----ArgGluGluProProLeuProProProProPr 2880
QY 1714 TCCAGGCGGAGCTGGCGC 1732
Db 2880 oProGlnProLeuTriPha 2886

RESULT 11
0905K9 PRELIMINARY; PRT; 608 AA.
AC 0905K9;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE NTR.
OS Herpesvirus papio.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BABOON LYMPHOCRYPTOVIRUS BA65;
RA Zong J.-C., Ryan J., Ling P.D., Loeb D.D., Pagano J.S., Hayward S.D.,
RA Hayward G.S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF200364; AAF23950.1;
SQ SEQUENCE 608 AA; 60458 MW; 4BF82ACB0A029627 CRC64;

Alignment Scores:
Pred. No.: 1,05e-08 Length: 608
Score: 332.00 Matches: 188
Percent Similarity: 32.71% Conservative: 22
Best Local Similarity: 29.28% Mismatches: 244
Query Match: 9.93% Indels: 188
DB: 12 Gaps: 35

US-10-009-782-1 (1-1758) x 0905K9 (1-608)
QY 1687 CCTGCGACAGGCTTCGACAGGCTCATTTTCCGGGCTTCGCGCGCGAGCGCGCGC 1628
Db 10 ProGlyProGlyArgProGlyThrPhProGlyArgGluAlaGlyArgGluAlaGlyArg 29
QY 1627 CCATCACTCTCTGCG-----CCAGATATCGGCTG----- 1598
Db 30 ThrTrpGlyProCysProAlaProSerProProProAlaProAlaValHisGlyProArg 49
QY 1597 GGGCGGTATTCGAGCGGAGGAGGAGG-----CCATGTAGC 1562
Db 50 SerProAlaGlyThrGluArgArgArgGlySerAlaGlnArgGlyHisProProProGlyAla 69
QY 1561 GGTTCGCTCGAGGCTTCGAGGCGGCGCGCGCGCGC-----CCC 1523
Db 70 GlyGlnArgProSerGlyProThrGlyGlyHisProAlaAlaProGlyAlaProGlyPro 89
QY 1522 CGTTCACGCGGATGTAG-----GCTGCGC 1496
Db 90 ArgSerProAlaGlyThrGluArgArgArgGlySerAlaGlnArgGlyHisProProProGly 109
QY 1495 GCCGGGCTCAGCGCGCGCTGCGTTCGACAGCGCGCGCGCATGCTGCGCGGTGAAGCGCT 1436
Db 110 AlaGly-----GlnArgProSerGlyProThrGlyGlyHisProAlaAlaProGlyAlaPro 128
QY 1435 GCTTCGCGACAGCGCGCGCGCTGTCAGCTACAGGATGATGCGCGCGCGCGCTCGG 1376
Db 129 GlyProArgSerPro-----ArgThrGluArgArgArgGlySerAlaGlnArgGly--- 145
QY 1375 TAGGCTGTCGAAGGTGGGGGTATCGGCACCGTGGCGCGGCTCGAAGCACACAGGCTCGG 1316
Db 146 -----HisProProProGlyAlaGlyGlnArgProSerGlyPro 158

```

```

QY 1315 CGAATACCCGCGCTTCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1256
Db 159 ThrGlyGlnHisProAlaAlaProGlyAlaProGlyProAlaGlySerProAlaGlyArg 178
QY 1255 CGGTCACTTTCGAA-----CCGCGCTTCGACGGGA-----ACAGCG 1217
Db 179 ArgArgGlySerAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGly 198
QY 1216 CCAGTCCGCGCGCAT-----AGTCCCGCA-----GCACCGCGGGAAGTCCCGCGCGCGCG 1163
Db 199 ProThrGlyGlyHisProAlaAlaProGlyAlaProGlyProAlaGlySerPro----- 215
QY 1162 GATCGCGCGCTCGCTGCGCGAGCGCTCGAGCGCGCATGCTGCGCGCGCGCGCGCGCG 1103
Db 216 -----ArgThrGluArgArgGlySerAlaGlnArgGlyHisProPro 230
QY 1102 GGATGCGCTGACGCTCGGCTTCGCTCATGATGATGATGCGCGCGCGCGCGCGCGCT 1043
Db 231 ProGlyAlaGlyGln-----ArgProSerGlyProThrGlyGlyHisProAlaAlaPro 248
QY 1042 CGGCGACACGCTGCTACTGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 983
Db 249 GlyAlaProGly-----ProArgSerProAlaGlyThrGluArgArgGly 263
QY 982 CGCTCACTTCGCGGAGAGGCTTCGACACAGGTGATGATGCTGCGCGCGCGCGCGCGCG 923
Db 264 SerAlaGlnArgGlyHisPro-----ProProGlyAla 274
QY 922 GGTCTGCTTCGACATGCTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 869
Db 275 Gly-----GlnArgProSerGlyProThrGlyGlyHisProAla--- 287
QY 868 CGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 809
Db 288 AlaProGlyAlaPro-----GlyProArgSerProAlaGlyThrGluArgArgGlySerAla 306
QY 808 TGGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 749
Db 307 GlnArgGlyHisProProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyHis 326
QY 748 TCGGAAGGTTTCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 689
Db 327 ProAlaAlaProGlyAlaProGlyProAlaGlySerProAlaGlyThrGluArgArgGlySer 346
QY 688 CGTAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 653
Db 347 AlaGlnArgGlyHisProProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGly 366
QY 652 -----CCTCATGATCTCTTCGCTGCGCGCGCGCGCGCGCG 620
Db 367 HisProAlaAlaProGlyAlaProGlyPro-----ArgSerPro 379
QY 619 CGGCGCGGTGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
Db 380 ArgThrGlnArgArgArgGlySerAlaGlnArgGlyHisProProProGlyAlaGlyGln 399
QY 559 -----CSATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 536
Db 400 ArgProSerGlyProThrGlyGlyHisProAlaAlaProGlyAlaProGlyProAlaGlySer 419
QY 535 CCGTCTGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
Db 420 ProAlaGlyThrGluArgArgArgGlySer-----AlaGlnArgGlyHis 433
QY 475 CCACATACAGCGCGCGCTGACGCGCG-----CCGCGCGCGCGCGCGCGCGCGCGCG 437
Db 434 ProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyHisProAlaAlaPro 453
QY 436 GCAAGCGCTCAAGTACGCGCA-----AGCGCTCGAAGC 401
Db 454 GlyAla-----ProGlyProArgSerProAlaGlyThrGluArgArgGlySerAlaGlnArg 472
QY 400 GGTAG-----AGCGCGCTTCGTCACAGAGTCA-----GGGCGCGCGCGCGCGGT 356

```

```

Db 473 GYHSPROFROGLYALAGLYGlnArgProSerGlyProThiGlyGlnHis----- 490
OY 355 TGGCGTGGCCGCGCGCGCGCGAGTGGCCGCAATGGCCGTCGACAGCCGTGGTGGCC 296
Db 491 -----ProAlaAlaProGly-----AlaProGlyProArgSerProArg 503
OY 295 CCTGGAGATCTGGGGCGGCGAGTGGCGAGCGCTGACAGGATGGTGGTGGTGGTGGT 236
Db 504 ThrglnArgArg-----ArgGlySerAlaGlnArgGlyGlnHis 515
OY 235 GCGAGTCGATGAGCGCGCGCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 188
Db 516 ProProProGlyAlaGlyGlnArgProSerGlyProThiGlyGlnHisProAlaAlaPro 535
OY 187 GGGTGT-----GCGCGCGCGCGCGCGCGCGAGTGGCGCGCGCGCGCGCGCGCGCG 131
Db 536 GYAlaProGlyProArgSerProArgThrglnArgArgGlySerAlaGlnArgGly 555
OY 130 CGCCCA-----GCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 110
Db 556 HisProProProGlyAlaGlyGlnArgProSerGlyProThiGlyGlnHisProAlaAla 575
OY 109 CCGGGG 104
Db 576 ProGly 577

RESULT 12
O80ZE1 PRELIMINARY; PRT; 890 AA.
AC O80ZE1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE L53.
OS cercopithecine herpesvirus 15.
OC Viruses; dsDNA viruses, no RNA stage: Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=104228;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RA MEDLINE=97048062; PubMed=8892903;
RX Franken M., Deyergne O., Rosenzweig M., Anis B., Kieff E., Wang F.;
RT "Comparative analysis identifies conserved tumor necrosis factor
RT receptor-associated factor 3 binding sites in the human and simian
RT Epstein-Barr virus oncogene LMP1."
RL J. Virol. 70:7819-7826(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RA MEDLINE=99412410; PubMed=10482645;
RX Rivallier P., Quink C., Wang F.;
RT "Strong selective pressure for evolution of an Epstein-Barr virus
RT LMP2B homologue in the rhesus Lymphocryptovirus."
RL J. Virol. 73:8867-8872(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RA MEDLINE=20304984; PubMed=10846073;
RX Jiang H., Cho Y.-G., Wang F.;
RT "Structural, functional, and genetic comparisons of Epstein-Barr virus
RT nuclear antigen 3A, 3B, and 3C homologues encoded by the rhesus
RT Lymphocryptovirus."
RL J. Virol. 74:5921-5932(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RX MEDLINE=20440633; PubMed=10970361;
RA Rao P., Jiang H., Wang F.;
RT "Cloning of the rhesus Lymphocryptovirus viral capsid antigen and
RT Epstein-Barr virus-encoded small RNA homologues and use in diagnosis
RT of acute and persistent infections."

```

```

RL J. Clin. Microbiol. 38:3219-3225(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RX MEDLINE=21602573; PubMed=11739708;
RA Rivallier P., Jiang H., Cho Y.-G., Quink C., Wang F.;
RT "Complete Nucleotide Sequence of the Rhesus Lymphocryptovirus: Genetic
RT Validation for an Epstein-Barr Virus Animal Model."
RL J. Virol. 76:421-426(2002).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RA Moghaddam A., Koch J., Anis B., Wang F.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RA Moghaddam A., Anis B., Wang F.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RA Rivallier P., Quink C., Wang F.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RA Jiang H., Wang F.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RA Rao P.V., Jiang H., Wang F.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL8664;
RA Rivallier P., Jiang H., Cho Y.-G., Quink C., Wang F.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY037858; AAK95470.1;
SQ SEQUENCE 890 AA; 95097 MW; 149077C89D3B5574 CRC64;

Alignment Scores:
Pred. No.: 1,75e-08 Length: 890
Score: 318.00 Matches: 200
Percent Similarity: 34.73% Conservative: 32
Best Local Similarity: 29.94% Mismatches: 217
Query Match: 9.81% Indels: 220
DB: 12 Gaps: 45

US-10-009-782-1 (1-1758) x O80ZE1 (1-890)
OY 1750 CGTAGCTCAGGCGGATCTGGCCGAGCTCGGCT-----TGACAGGCGTGA 1703
Db 51 ArgGlySerGlyProArgGlyThiArgProProAlaProSerSerArgThiGlyGlnArg 70
OY 1702 CCGCGATGCGCGCTGGCGCTGGACAGGTCTTGACAGGTCATTTCCGCGGCTTCGCTCGG 1643
Db 71 ThArgGlnArg-----ProGlyHisProProProGlyArgGlySerGlyPro----- 86
OY 1642 CCGCGAGCGCGCGCGCGCGACACTCTCTGCCACGATACGGGTGGGG----- 1595
Db 87 -----ArgGlyThiArgPro-----ProAlaProSerSerArgThiGlyGlnArgThiArg 103
OY 1594 -----CCGATATTCGAGCGGAGGAGGAGCGCATGTAGCGGTTGGTGGTGA 1550
Db 104 GlnArgProGlyHisProProProGlyArgGlySerGlyPro-----ArgGlyThiArgPro 122
OY 1549 GGGTGGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1496
Db 123 ProAlaProSerSerArgThiGlyGlnArgGln-----ArgProGlyHisPro 140
OY 1495 -----GCCGGGCTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1460

```

```

Db 141 ProProGluArgGlySerGlyProArgGlyThrArgProProAlaProSerSerArgThr 160
QY 1459 CGGATGCTGGCCCGGTGACACCGCTCTTGGCCAGACGGCGCGC----- 1415
Db 161 GlyAlaThrArgGlyAlaArgGlyAlaProProGluArgGlySerGlyProArg 180
QY 1414 -----CGTTGACGATACAGGAAATGGATCCGCGCGCGCGCGCGTGGT 1370
Db 181 GlyThrArgProProAlaProSerSerArgThrGlyAlaThrArgGlyAlaArg 198
QY 1369 GTTCGAAGAGTGGCGGATATCGGACCGGTGGCTGACACACCA----- 1322
Db 199 -----ProGlyHisProProGluArgGlySer 208
QY 1321 GGTGCGGAGATACCGCGCTGACAGTCCGCGCGCGCGCGCGCGCGCGCGCG 1265
Db 209 GlyProArgGlyThrArgProProAlaProSerSerArgThrGlyAlaThrArgGln 228
QY 1264 CGGTGACGCGCGGTGATCTCCATACCGCGCTGACAGCGGAGCGCGCGCGCG 1205
Db 229 ArgProGlyHis-----ProProProGluArgGlySerGlyPro-----ArgGly 243
QY 1204 CATAGTCCCGACAGCCCGCGGAGAGTGGCCACAGCGCGGATGCGCGCTGCTGT 1145
Db 244 ThrArgProProAlaProSerSerArg----- 252
QY 1144 GCGGACGCGCGTGGACGATCATGTCGCGCGCGCGCGCGCGCGCGCGCGCG 1094
Db 253 -----ThrGlyAlaThrArgGlyAlaArgProGlyHisProProGluArgGlySerGly 271
QY 1093 GCACGTGCGGCTGCTGCTCATGATGAGATGAGCGCGCGCGCGCGCGCGCG 1037
Db 272 ProArgGlyThrArgPro-----ProAlaProSerSerArgThr 284
QY 1036 CCACGCTGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 977
Db 285 GlyAlaThrArg-----ArgGlyAlaArgProGly-----HisProProGluArgGly 300
QY 976 GTTCGCGGAGAGGCTGACACAGGTATGATGCTGCGCGCGCGCGCGCGCGCG 917
Db 301 SerGlyProArgGly-----ThrArgProProAla-----Pro 311
QY 916 GCTTGAGCATGTGAGCCGCGCACGTAGGAGATACGCGTCCAGGACGACGCTG 857
Db 312 Ser-----SerArgThrGlyAlaArgThrArg-----GlnArgProGlyHis 325
QY 856 CCATGCGCGCGCTGATCAGCGCGAGCTGCGCGCGAGCGCGCGCGCGCGCG 797
Db 326 Pro-----ProProGluArgGlySerGly----- 333
QY 796 TGACCTTGCTGTCGAGATCACACCGCGCA-----CGTCCAGCTGCG 755
Db 334 ---ProArgGlyThrArg-----ProProAlaProSerSerArgThrGlyAlaThrArg 351
QY 754 GGCCTGATGCGGAGGTTCTCTCCAGCGCG-----CCAGATGTCGCTGCTG 704
Db 352 GlnArgProGlyHisProProProGluArgGlySerGlyProArgGlyThrArg---Pro 370
QY 703 CGCGCATGTGGGTGCGTAGATGCCCGCATGCGCGTCCAGCGCGCGCGCGCGCG 644
Db 371 ProAlaProSerSerArgThrGlyAlaThrArg-GlnArgProGlyHisProPro 390
QY 643 TCTCTGCTGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 591
Db 390 oGluArgGlySerGlyProArgGlyThrArgProProAlaProSerSerArgThrGly 410
QY 590 CGCATGCGCGCGCGCATGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 543
Db 410 nArgThrArgGlnArgProGlyHisProProProGluArgGlySerGlyProArgGly 430
QY 542 GCGATTTCT-----CGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCG 504

```

```

Db 430 rArgProProAlaProSerSerArgThrGlyAlaThrArgGlnArgProGlyHisPr 450
QY 503 ATGACCGCGCGCGCGAGCGTTGATGATGCGCCACCATACAGAGCGCGCTGACGCGCGCG 444
Db 450 oProProGluArgGly-----SerGlyPro-----ArgGlyThrArgProAla 465
QY 443 GTGCGCGCGCGCGCGCTCCAGGTAGTCGGCGAAGCGCTGCA----- 404
Db 465 a---ProSerSerArgThrGlyAlaThrArgGlnArgThrArgGlnArgProGlyHisProPro 484
QY 403 -----AACGCTAGACCGCGCTGCTGTCACAGAGTCCAGCGCGCGCGCGCGTGGCG 351
Db 484 uArgGlySerGlyProArgGlyThrArgProProAlaProSerSerArgThrGlyAlaThr 504
QY 350 TGCGCGACGCGCGCGCGCGCTGATGCGCGCAATGCGCGTCCAGACCGCGTGGTACGCGCTG 291
Db 504 gThrArgGlnArgProGly-----HisProPro-----ProGly 515
QY 290 GAGATCTTGGCGGCTGATCTGCGGACGCTGACAGAGTATGTCGTGCGGTGTCGAG 231
Db 515 uArg-----GlySerGlyProArgGly----- 522
QY 230 TCGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 171
Db 523 -----ThrArgProProAlaProSerSerArgThrGlyAlaThrArgGly 538
QY 170 T----- 170
Db 538 nArgProGlyHisProProProGluArgGlySerGlyProArgGlyThrArgProProAla 558
QY 169 -CCGACAGATGCGCGAGTGGCGCGATGCGCGCGCGCGCGCGCGCGCGCGCG 122
Db 558 aProSerSerArgThrGlyAlaThrArg-----ThrArgGlnArgProGlyHisProPro 576
QY 121 -----CGGCGCGCGCGCGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 66
Db 576 oGluArgGlySerGlyProArg-----GlyThrArgProProAlaPr 590
QY 65 AGTTCGAAGCGCGTGGATCGG 44
Db 590 oSerSerArgThrGlyAlaThrArg 597

```

RESULT 13

ID 052747

AC 052747

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Endo-glucanase.

OS Ruminococcus flavefaciens.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Lachnospiraceae; Ruminococcus.

OX NCBI_TaxID=1265;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=186;

RA Huang C.M., Asmundson R.V., Yu P.L.;

RT "Nucleotide sequence of a cellulase gene complex from Ruminococcus

RL flavefaciens strain 186 coding for multi-cellulase activities.";

DR Submitted (SEP-1990) to the EMBL/Genbank/DBD databases.

DR EMBL: X56082; CAA39559.1;

SQ SEQUENCE 680 AA; 75503 MW; 5A589F334846D6F CRC64;

Alignment Scores:

Pred. No.: 5 05e-08

Score: 307.00

Percent Similarity: 33.24%

Best Local Similarity: 28.01%

Query Match: 9.31%

DB: 2

Gaps: 45

Length: 680

Matches: 198

Conservative: 37

Mismatches: 195

Indels: 277

Gaps: 45

US-10-009-782-1 (1-1758) x 052747 (1-680)

```

QY 32 CCAATGCGCAATCGGATT-----CCAGCCCTTGAGACTGCTGCTCCGGGCG 79
DB 20 ProSerProSerProThSerProThProProSerProSerTh-----GluSer 36
QY 80 GCACCCCTA---TCAGCGGCAACAACACCCGGGGCGGCGGACCTGGGCGGCGG 136
DB 37 SerProThProSerSerProAlaSerProArg-----SerProThValArgGlyArg 54
QY 137 GCGAGCCGATGCGCGGCAATCGGCGATCTGTGAGACCGCGCGGCGACACCGGGTGCAG 196
DB 55 GluAlaValAlaPro-----AlaGlySerAlaSerAlaArgThr--AlaGlyArg-- 70
QY 197 TGTGGGCGCTGTGTGTGCGCGCGCGGCTTCATGCACTGCGCAACCCAGACACAACTACC 256
DB 71 -----GlyArg-----AspProArgHisGlnCysAs 79
QY 257 TGTCTAGGCGCTGCGGACATGACGCCAAGATCTGCGAGGGCGGTGCACACAGCT-----GG 310
DB 79 PAlaArg-----LeuArgGlyArgHisArgProArgGlyLeuG 91
QY 311 TCAGCGGCAATTCGG-----326
DB 91 YHisAlaValAlaArgAspArgGlyLeuGlyProHisGlnArgAspCysValAlaArgGlyAl 111
QY 327 -----CATCAGCGCTGGCGCGCTGGCGGACG 352
DB 111 AcGlyHisGlyProSerValHisArgAlaSerThrHisGlnProArgArgAlaAlaLeuG 131
QY 353 C-----CAACC 358
DB 131 ULeuValGlyLeuValHisProGlnSerGlyLeuValProValAlaArgGlyValPr 151
QY 359 CGCGCGCGCGCTGAGCTGTGAGAGAGCGGCTTACCGTTTGAGACGCTTGGCGG 418
DB 151 AlaArgProArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 171
QY 419 ACTACCT---GAGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGTGTGTG 475
DB 171 nLeuProLeuProArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 185
QY 476 GGCATTCAGCGCTGCGGCGGCGGCGGCTCATGCCGACTTGACAGCGCGCGCGCGGCGG 535
DB 186 -ProHisAspAla-----189
QY 536 AATTCGCGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTT 595
DB 190 -AspAlaValHisGlyGly---GlyArgGlyValArgGlyGln-----202
QY 596 CGACGCGCGCTTCTACCGCGCGCGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGG 652
DB 203 -----LeuGlnSerArgArgArgArgArgArgArgArgArgArgArgArgArgArg 218
QY 653 TGTGGCGGCG-----GCTGAGCGGCGCATG 676
DB 218 pGlyProSerAspArgArgArgSerPheAspProProAspArgArgAlaValArgArgG 238
QY 677 GCGGCTTACCGGCGAC-----CC 694
DB 238 nArgGlyArgArgHisArgLeuValAspArgArgValArgProArgGlnGlnProValPr 258
QY 695 ACATGCGCGGACGAGG-----CGAGCAGCATGCTGGCGCGGCTGAGGAAACT 742
DB 258 oArgValAlaArgGlnArgGlyProGlyValArgGlnHisGln-----271
QY 743 TCCGCGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 802
DB 272 -----HisArgArgArgArgGlyProLeuAlaGlyAspLeuValGlnArgAlaGlnAla-- 289
QY 803 AGCCCAATTCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 862
DB 290 -AlaHisProArgSerValGlnArgGlnProIle-----AspHisGly----- 303

```

```

QY 863 AGGAGCTGTGCTGAGCGGCTGATCCCTACGTGGCGCGGCGGCGGCGGCGGCGGCGG 922
DB 304 -GlyArgGlyLeuHisAlaArgValGlyAlaGlnGlnArgGlyAlaArgGlyAspArgGly 323
QY 923 GCGTCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 982
DB 323 P---GlyAlaArgGlyArg-----327
QY 983 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1042
DB 328 -----GluArgGlyArgAlaGlnArgGlyLeu-----ArgArgGlyGlyArg 340
QY 1043 A-----GTCAGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1072
DB 340 gLeuHisArgArgGlnArgGlnArgGlnAlaSerAlaValGlnGlnArgHisProGlyArgArg 360
QY 1073 TGTAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1112
DB 360 oSerGlyLeuGlnArgArgGlnAlaSerAlaValGlnGlnArgHisProGlyArgArgArg 380
QY 1113 -----CCGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1162
DB 380 gProProGlnGlnArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 400
QY 1163 CGCGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1222
DB 400 aGlyAlaSerGlyGlnArgArgValGly-----409
QY 1223 TCCCGGTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1282
DB 410 -ProArgGlyGlyGly-----ValG 416
QY 1283 CGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1328
DB 416 YArgLeuArgValSerAlaLeuArgValAlaProLeuArgArgArgArgArgArgArgArg 436
QY 1329 -----GTTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1375
DB 436 gAlaAspCysHisAlaArgProGly-----ValSerHisArgAlaThrPro-- 451
QY 1376 CCGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1409
DB 452 -----ArgGlyHisProHisArgCysArgAlaValAlaArgGlySerArgGlnAspArg 468
QY 1410 -----CAACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1453
DB 468 pArgAlaArgAspLeuArgArgArgHisValAspAlaThrAlaGlyGlnGlyArg--Ar 487
QY 1454 ATCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1507
DB 487 gAspArgProGlnLeuArgValAlaValHis-----G 498
QY 1508 AATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1564
DB 498 nLeuArgArgGlnAlaThrGlyGlnArgProAspAlaMetGlyAlaArgAlaHisProIl 518
QY 1565 ACATGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1624
DB 518 eHisThrGlyGlyValArgLeuProArgGlyAlaArgAlaSerGlnGlyHisArgArgAs 538
QY 1625 TGGGCGCGGCG-----CCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1672
DB 538 pArgThrAlaValAlaArgLeuProArgValAlaArgArgAspSerArgAlaArgArgThrAr 558
QY 1673 AAGACCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1732
DB 558 gAlaAspValVal-----ArgAlaArgArgLeuArgThrArgArgGlyAlaAlaGlyHis 576
QY 1733 AGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1751
DB 576 rGlnArgAlaGlnLeuArg 582

```

RESULT 14

Db	133	-----	ATGAspGlyValAsnArgGlyAsp-----	GlyAlaProGluArgGlySerGly	148
QY	1348	CCACCGTGGCCGGGTGCGAACACACAGAGTCCGGCGAAGTACCCGGCTGCAGCTGCCCC	128		
Db	149	LeuProGlyProGlyCysGlySerProGlyLeuGlySerArgSerProProAlaGlyTyr	168		
QY	1288	GCCCGCCAGGC-----	CGAAGCGCGCGGGTCAAGCCGG	125	
Db	169	GlyGluProGlyGlyValArgGlyProGlyAlaArgGluAlaAlaArgProGlyValAlaArg	188		
QY	1252	-----	TCAATCTTCATACCCGCG-----	TCTCCACCGCGAACAGCC	121
Db	189	ValProAlaGlyProGlyGluProAlaProAlaProGlyGlyMetGlyGlyThreGlyGlyThreGly	208		
QY	1216	CCAGGTCCGCGGCATAGTACCCCGCACACCCGCG-----	GGAAAGTGGCCCCACAGCGCGGAT	116	
Db	209	Pro-----	ArgSerGluGlyValAlaGlySerProAlaAlaArgGlyAlaGlyProProAlaTyrGly	227	
QY	1159	GCGGCG-----	GCTCGTCGTGCGCGGACAGCCGTCGG-----	113	
Db	228	AlaGlyProProAlaGlyArgAspGlyValGlyAsnArgGlyAlaAspGlyAlaProGluArgGlySer	247		
QY	1129	-----	AGCCGATCA-----	112	
Db	248	GlyLeuProGlyProGlyCysGlySerProGlyLeuGlySerArgSerProProAlaGly	267		
QY	1120	TGG-----	TCGGCGCGAAGCCAGAGATGCGTGCACGTCGGGTTCGTCCATCATGA	107	
Db	268	TyrGlyLeuProGlyGlyArgGlyProGlyAlaArgGluArgAlaAlaArgPro-----	285		
QY	1069	AGTAGATGGCGCGCGCGGTGCGAGCTCGGGCACACAGCTCTACTGATTTGCCGCGCT	101		
Db	286	-----	GlyValArgValProArg-----	291	
QY	1009	CGGCGCGAGCTCATCCAGTCCG-----	111111	986	
Db	292	-----	ProGlyGluProValProProGlyGlyMetGlyGlyThreGlyGly	306	
QY	985	-----	GCCGCGCATGTTCCGGGAGAGGGCTTCGACACAGTATATAGTGGCTCCGGCCA	932	
Db	307	ThreGlyProArgSer-----	GluGlyAlaGlyCys-----	Pro	317
QY	931	GCAACAGCGCGTCTCGTGTAGCATGTGTGAGCGCGCACAGTACAGGATACGCGTCAACG	872		
Db	318	AlaArgGlyValAlaGlyProProAlaTyrPheGlyAlaGlyProProAlaGlyAspGlyValAsn	337		
QY	871	AGACGCTCTGGCGCGCCATGCGCGCTCGATCAACGCGCAGCTCTCGCGAGCGCGCA	812		
Db	338	ArgGlyAspGlyValaProGluArg-----	GlySer	347	
QY	811	AATTGGGCTGGCCCATGACCTGTT-----	GTCGCGAGATACACACCGCGCA	767	
Db	348	GlyLeuProGlyPro-----	GlyCysGlySerProGlyLeuGlySerArgSerProProAla	366	
QY	766	-----	CGTCCAGCTCCGCGCGCCATGCGGAGAGTTCTCTCCAGCGCGGCCACAGTGT	716	
Db	367	GlyTyrGlyGluProGly-----	GlyArgGlyProGlyAlaArgGluArgAla	382	
QY	715	GCTCGCCTTCGTGCGCATGTGGGTGGCTAGATGCCCGCCATGCGCGCTCAGCGCGCGC	656		
Db	383	AlaArgProGly-ValaArgVal-----	ProArgProGlyGluProVal	396	
QY	655	ACACCTCGATGATCTTGGTGGTGGCGCGCGCGCGCGGGCGGGTGAAGCGCCGCGTCG	596		
Db	396	LeuProGlyGlyMetGlyGlyThreGlyGlyThreGlyProArgSerGluGlyAlaGlyCys	416		
QY	595	AAATCCGATGGCGCGCGCTGGCCATGCGTCTCTCGCGCAGTCCCGCATATGCGCGCATTT	536		
Db	416	SerProAlaArgGly-----	AlaGlyProProAlaTyrPheGlyAlaGlyGlyProProAlaArgAsp--	434	
QY	535	CCTGTCGTGGCGCGCGCTGCAAGTCCGCGCATACCGCGCGCGCAGCGTTGATGGC	476		


```
Db 435 -----GlyGlyAsnArgGlyAspGlyAlaProGluArgGly--SerGlyLeuProGly 451
QY 475 CCA-----CCATACAGCGCGCGTGTGACGCGCGCGG---TGG-----CCC 437
Db 452 ProGlyCysGlySerProGlyLeuGlySerArgSerProAlaGlyTrpGlyGluPro 471
QY 436 GCAAGCGGT---CCAGGTAGTGG---CGAAGCGGTGGAAGCGTGAAGACCGCCTTGT 383
Db 472 GlyGlyArgGlyProGlyAlaArgGluArgAlaAlaArgProGlyAlaArgValProArg 491
QY 382 CCAGCAGGTCCA-----GGGGGCGGGGGGTTGGCGTGGCGCCAGCG 341
Db 492 ProGlyGluProValProProGlyGlyMetGlyGlyThrGlyGly---ThrGlyProArg 510
QY 340 GCGCCAGGCTGATGCCGCAATTGCCG-----TGACCACCGGTGGTGAAGCCCT 293
Db 511 SerGlyGly-----AlaGlyCysProAlaArgGlyAlaGlyProPro----- 524
QY 292 GCGAGATCTTGGCGCATGTGCGGACGCGCTGACGAGTGTGTGTGCT 242
Db 525 -----AlaTrpGly-----AlaGlyProProArgArgAspGlyGly 536
QY 241 -----GGGTGTGCGAGT 230
Db 537 AsnArgGlyAspGlyAlaProGluArgGlySerGlyLeuProGlyProGlyCysGlySer 556
QY 229 CGATGAAGCGCGCGCGCA---CCACCAAGCGCCGACACGTGACCGCGGTGGCGCGCG 173
Db 557 ProGlyLeuGlySerArgSerProAlaGlyTrpGlyGluProGlyGlyArgGlyPro 576
QY 172 CGTCCGACAGATGCCGCGATGCGCGATGCGCGGTCGCGCCAGCGCGTGGCGCGCC 113
Db 577 GlyAlaArgGluArgAlaAlaArgProGlyValArgValProAlaArgProGlyGluProVal 596
QY 112 GCCCGGCGGTGTGTCGCGTGCATGAGGGTGGCGCCCGCGACAGCAGGT----- 62
Db 597 ProProGlyGlyMetGlyGlyThrGlyGlyThrGlyProArgSerGlyGlyAlaGlyCys 616
QY 61 -----CGAAGGCTGGGATCGGATGGGACATGGAATCTCTCCCTCCGCGATCAAGT 8
Db 617 ProAlaArgGlyAlaGlyPro-----ProAlaTrpGlyAlaGlyProProArgArgAsp 634
QY 7 GGA 5
Db 635 Gly 635
```

Search completed: May 11, 2003, 12:15:53
Job time : 113 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 01:33:19 ; Search time 82 Seconds

(without alignments)

6574.851 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 1758

Sequence: 1 gaattccactgacgcgcga.....ccctgagctacgagaagctt 1758

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/PCRTS.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	94.2	5.4	US-07-945-283-1	Sequence 1, Appl
2	81.8	4.7	US-09-105-537-7	Sequence 7, Appl
3	81.8	4.7	US-09-105-537-3	Sequence 3, Appl
4	77.4	4.4	US-09-773-816-1	Sequence 1, Appl
5	76.2	4.3	US-09-320-878-21	Sequence 21, Appl
6	75.6	4.3	US-08-804-227C-1	Sequence 1, Appl
7	75	4.3	US-08-461-775-8	Sequence 8, Appl
8	75	4.3	US-09-031-606-8	Sequence 10, Appl
9	75	4.3	US-08-461-775-10	Sequence 10, Appl
10	75	4.3	US-09-031-606-10	Sequence 10, Appl
11	74.4	4.2	US-08-458-568A-11	Sequence 11, Appl
12	73.8	4.2	US-09-105-537-1	Sequence 1, Appl
13	73.4	4.2	US-08-461-775-9	Sequence 9, Appl
14	73.4	4.2	US-09-031-606-9	Sequence 9, Appl
15	73.4	4.2	US-08-461-775-11	Sequence 11, Appl
16	73.4	4.2	US-09-031-606-11	Sequence 11, Appl
17	73.4	4.2	US-08-690-473-1	Sequence 1, Appl
18	73.4	4.2	US-09-259-821A-1	Sequence 1, Appl
19	73.4	4.2	US-08-843-659-1	Sequence 1, Appl
20	72.8	4.1	US-09-385-028-13	Sequence 13, Appl
21	72.8	4.1	US-09-385-028-1	Sequence 1, Appl
22	72.2	4.1	US-09-105-537-30	Sequence 30, Appl
23	72.2	4.1	US-09-105-537-5	Sequence 5, Appl
24	72.2	4.1	US-09-320-878-19	Sequence 19, Appl
25	71.2	4.1	US-08-403-852D-6	Sequence 6, Appl
26	71.2	4.1	US-08-510-646B-6	Sequence 6, Appl
27	71.2	4.1	US-09-231-818-6	Sequence 6, Appl

28	71.2	4.1	2634	3	US-08-941-936-1	Sequence 1, Appl
29	70	4.0	4403765	4	US-09-103-840A-2	Sequence 2, Appl
30	69.6	4.0	11220	4	US-09-105-537-32	Sequence 32, Appl
31	68.8	3.9	5392	2	US-08-403-852D-1	Sequence 1, Appl
32	68.8	3.9	5392	2	US-08-510-646B-1	Sequence 1, Appl
33	68.8	3.9	5392	4	US-09-231-818-1	Sequence 1, Appl
34	68.6	3.9	44377	2	US-08-804-227C-7	Sequence 7, Appl
35	68.6	3.9	44377	2	US-08-804-198-1	Sequence 1, Appl
36	68.2	3.9	2220	4	US-08-765-907A-14	Sequence 14, Appl
37	68.2	3.9	4496	4	US-08-765-907A-6	Sequence 6, Appl
38	68	3.9	2064	1	US-08-343-428-1	Sequence 1, Appl
39	68	3.9	12588	2	US-08-387-942C-1	Sequence 1, Appl
40	67.8	3.9	1014	4	US-09-105-537-13	Sequence 13, Appl
41	67.4	3.8	30001	1	US-08-125-468-1	Sequence 1, Appl
42	67.4	3.8	30001	2	US-08-474-933-1	Sequence 1, Appl
43	67.2	3.8	1208	2	US-08-403-852D-4	Sequence 4, Appl
44	67.2	3.8	1208	3	US-08-510-646B-4	Sequence 4, Appl
45	67.2	3.8	1208	4	US-09-231-818-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-07-945-283-1/C
Sequence 1, Application US/07945283

Patent No. 5352596

GENERAL INFORMATION:

APPLICANT: Cheung, Andrew K.

APPLICANT: Wesley, Ronald D.

TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants

TITLE OF INVENTION: Involving The EPO and LIT Genes

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Curtis P. Ribando

STREET: 1815 No. 5352596th University Street

CITY: Peoria

STATE: IL

COUNTRY: USA

ZIP: 61604

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/945,283

FILING DATE: 19920911

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Ribando, Curtis P

REGISTRATION NUMBER: 27976

TELECOMMUNICATION INFORMATION:

TELEPHONE: 309-685-4011 ext. 513

TELEFAX: 309-685-4128

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 8438 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Pseudorabies virus

FEATURE:

NAME/KEY: CDS

LOCATION: 622..6495

FEATURE:

NAME/KEY: variation

LOCATION: replace(1099, "g")

	Query Match	Similarity	4.7%;	Score 81.8;	DB 4;	Length 1248;
	Best Local	Similarity	44.2%;	Pred. 1.6e-07;		
	Matches	439;	Conservative	0;	Mismatches 542;	Indels 12;
					Gaps	2
QY	32	CCATATCCCAATCCGATGTTCCAGCCCTTGGACTGTGCTGCGGGGGGACCCCTATCG	91			
Db	215	CCACCTGGACGCCACAGCGCGGGCTTCAGCTCTCGCGACGCGCGGCTTCACCGGGG	274			
QY	92	ACGGACGACACACCCCGGGGGCGCGCCGACCTGGCGCTGGCGGGACCCGACATCGCG	151			
Db	275	AAGTATCATGCGCTGATGATGACGTTCGCGCGCACCCCGGACGACATCGCCTTGATCGCC	334			
QY	152	CCATGGGGATCTGTCGGAGCGCGCGCGGACACCCGGGTGCACTGTGGCGCTGTGG	211			
Db	335	TCACCCCGGTTCGCGGACATCGACCCCGGACCGGACACTTCACCCGACCGACGGTGG	394			

OY	212	TGCGCCCGGGCTTATGCACTCGGACACACCACAGACAACTACCTGCTAGCGCTGCGG	271
Db	395	CGCGCGGGGTCAACACCCCGCACTCGGGCGTGTGGCGGTCCACTCTCTGGGGCGGCCCT	454
OY	272	ACATGACGCCCAAGATCTCGGAGGGGCGCTACACAGTGTGTACAGGGCAATTGGGGCATCA	331
Db	455	GGCGCGCGGACACAGCTGGGGAAAGGTCCGCGACAGACAGCGGCTTCGGCTTACTTTGACG	514
OY	332	GCCTGGCCCGCTGGCGGACGCCAACCCGCGCCCGCCCTTGAGACTTCGTGAGAAAGCG	391
Db	515	CGGCGACAGCCCTCGGGCTGCGGTCCACAGGCGCGGCGCCGCGGAGCGCTCGGCGAGCGCG	574
OY	392	GCTCTTACCGTTTGGAGGGCGTTCGCGCATCTACAGACGGGTGTGGGGCCACGCCCGCGG	451
Db	575	AGGCTTTAGCTTCCA---GCCACACAGGCGCTTCAACGCTTTCGAGGGGGCGCGCTG	631
OY	452	CGGTAAACGCGCGCTGTATGTGTGGGCCATTCAAGCTGCGCGCGCGGGGTATCGCGACT	511
Db	632	TCACCGAGGACGCGGACCTCGCGCGCGGATTCGCGCGCTTCACAACTTGGCTTGACCA	691
OY	512	TGCAAGCGGCGGCCACGACGAGAAATCGCGGCATGCGGAACTTGGCCGAGAAAGCA	571
Db	692	TGCCCCGGGGAGGCCGCGCGCGGACCAACGCAAGATGAGACAGGCGCGCGCGCA	751
OY	572	TGGCCAGCGGGCGCATCGGCACTTTCGACCGGCGCGCTTCTACCGCGCGCGCGCGCGCA	631
Db	752	TGGGGCTTACCTCCCTGACGCGCTTTCGAGGTCACTGACCGGAAACGGCGCAACCAAG	811
OY	632	CCACCGAAGAGATCATCGAGGTGTGCGGCGCGGTGAGCGCGCACTGGCGGCACTTACGCA	691
Db	812	CGGGCTACCGGAGACACTCGGGGACCTCCCGGCGTCTGTCGCGCAACAGACGCGC	871
OY	692	CCCACTAGCGCGGAGAGAGGAGACACTCTTGGCGCGCGCTGGAGGAAACCTTCGCATCG	751
Db	872	ACGGCTTCACACACACCACTAGTACTGTATCTTCAGATCGACGAGGGCACACACCGGCTCC	931
OY	752	GCCCGAGCTGGAAGTCCCGGTGTGATCTCGACACACAAGATGATGGGCGCAAGCCAA	811
Db	932	ACCGGACACTGTGATGGAAGTCTCTGAAAGGCGAAGCGGTGACACACCGCGCTACTTCT	991
OY	812	TGCGGCGCTCGCGGAGAACGCTGCGCTGATCGAG-----GCCGCACTGGCGCGC	862
Db	992	CGCGGGCTGCAAGACTGGAAGCGGTACCGCGCGGAGCGCGACGCGCCCGCTGGCGACA	1051
OY	863	AGGACGTCTCGCTGGAAGCGGTATTCCTACGTGGGCGCGGTCCACATGTCACAGAGAAC	922
Db	1052	CCGAACGCTTCGCGCGCGGTGTGTCTCTGCTGCGGACCGGACCGCCATCGCGAGAG	1111
OY	923	GCGTGTCTGGCGCGGACGACCATCATCACTGTTGTCAAGCCCTTCCCGCACTGAGCG	982
Db	1112	ACATTCGCCCGGTGTCGCACTCTGTGGTCTGTGGCGGACCGCGGCGGCAACTGACG	1171
OY	983	GGCGCGACCTGGAAGAAGTGGCGCGGAGCGCG	1015
Db	1172	CGCGCACCGGAGACGGCGCCCGCGCGGTG	1204

RESULT 3
 US-09-105-537-3/C
 Sequence 3, Application US/09105537A
 Patent No. 6265202
 GENERAL INFORMATION:
 APPLICANT: Sherman, D. H.
 APPLICANT: Liu, H.
 APPLICANT: Xue, Y.
 APPLICANT: Zhao, L.
 TITLE OF INVENTION: DNA encoding methymycin and plikromycin
 FILE REFERENCE: 600.438US1
 CURRENT APPLICATION NUMBER: US/09/105,537A
 CURRENT FILING DATE: 1998-06-26
 NUMBER OF SEQ ID NOS: 43
 SOFTWARE: FastSeq for Windows Version 3.0

```

; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-3

Query Match      4.7%; Score 81.8; DB 4; Length 13613;
Best Local Similarity 44.2%; Pred. No. 2,1e-07;
Matches 433; Conservative 0; Mismatches 542; Indels 12; Gaps

QY      32  CCATGTCCCATCCGATTCACAGCCCTTGACCTGTGCTCGCGGGGACCCCTCATGCG 91
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12659  CCACCTGGAACGCCAGCGCGCGGCTCTACACTCTCTGCGACGCGCGGCTCACGGCG 12611
QY      92  ACGGACGCAACACCCCGGGGCGGCGCGGACCTGAGGCGTGCAGCGGCACCCGATCGCG 151
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12609  AAGGATCATAGCCGTGATGATGACCTTGCCGCCACCCCGACGACACTGCGGTGATGGCC 12555
QY      152  CCATCGGGGCAATGTGTGGAGACCGCGCGCCACACCCGGGTGAGGTGTGCGGCTGTGG 211
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12549  TCACCCCGGCTTCTGCGCGAATGACACCCGGAACCGGCAACTCTGACCCGGACCAAGGTGG 12499
QY      212  TCGGCGCGGCTTCATATGACATGACACACCCAGAGAGACAACTGCTACAGGGGTGCG 271
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12489  CCGCGCGGGGTCAACACCCCGACCTTGCGGCGGTGTGTGGGTCTACCTTGCGGGGCGGCGCT 12430
QY      272  ACATGACGCCCAAGATCTGCGAAGGCGGTCCACACCGGTGTACAGGGCAATTCGCGCATCA 331
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12429  GCGCGCGGACACAGCTGGCGGAAGGTGGCGGACAGACAGCGCTGGGTGTACTTGTAGACG 12370
QY      332  GCGTGGCGCGGCTGTGGCGCACAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12369  CCGCGCACGCGCTCTGCGGTGCGGTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12310
QY      392  GCTTTACGCTTTCAGAGCGGTTCGCGACATCTGACATCGAGCGGTGGCGGCAACGCGCGCG 451
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12309  AGGCTTTCAGCTTCCA-----CGCCACCAAGCGCTTAAAGCGCTTGTAGAGGGCGGCGCGTGC 12255
QY      452  CCGTCAACGCGCGCTGTATGATGTGGGCAATTCAACGCTCGCGCGCGGCTATCGCGGACT 511
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12252  TCAACGCGAGCGGACGACGCTGGCGGCGCGGATCCGCGCGCTCAACAACTTGGGCTGTGACG 12193
QY      512  TGCAGCGCGCGCGCGCACACGAGAGAAATGCGCGCATCGCGGACCTTGGCCGAGAGAACCA 571
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12192  TGCCTGGCGGCAACCCCGCGCGGAGGACCAACGCCAAMATGAGGAGAGCGCGCGCGCA 12133
QY      572  TGGCAGACGCGCGCATCGGCAATTTGACAGCGGCGCTTACCGCGCGCGCGCGCGCGCGCA 631
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12132  TGGGCTTCACCTTCTCTGACGCGGTTTCCGAGGTATATGACCGGACCGCGCGCAACGACG 12073
QY      632  CCAACGAGAGATCATGAGTGTCTCCGCGCGCTGAGCGCGCATGCGCGGCGCATACGCA 691
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12072  CCGGCTTACCGCGAGACGCTCGGCGGACGCTCCCGGGGTCTGTGTGGCGGACAGACGGCG 12013
QY      692  CCAACATGCGCGACGAGGCGAGCACTGTGGCGCGCGCTGAGAGAAACTTCCGCGATG 751
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12012  ACGGCTTCAACAACACACAGTACGTGATGCTGAGATGACAGAGGACCAACACGCGGATCC 11953
QY      752  GCGCGAGCGGAGCGCGCGCGGGGTGATCTGCAACACAAAGTATATGGGCGACCCCAATT 811
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11952  ACGGCGAGCTCTGTCATGAGAGGTCTGGAAGCGCGAAGGCGGTGCACACCGCGCTTACTTCT 11893
QY      812  TCGGCGCGCTCGCGAGAGAGCGTGCCTGTATCGAG-----GCGCGCATGGCGCGCG 862
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11892  CCGCGGGGTGCCAGAGGTGTGAGACCGGTACCGCGGCGACCGGACGCGCGCTCCGCCACA 11833
QY      863  AGGACGCTCTGCGTGAAGCGGTATCCCTTACGTGGCGCGGCTCCACACTGCTTCAGACGAGCC 922
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11832  CCGAAGCGCTTCGCGCGCGCGGCGGTGTCTCTTCGCGACCGGACCGGACGCGGTGAGAGAG 11773
QY      923  GCGTGTGTGTGCGCGGAGCGACCAATCATACACTGTGTGAGAACCGCTTCCCGAATGTAGCG 982
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11772  ACATTCGCGCGGCTGCGCGACGCTGTGCTCTCTCTGCGAGCGCGCGCGCGGCGGAACTGACG 11713

```


RESULT 5
US-09-320-878-21/c
Sequence 21, Application US/09320878A
Patent No. 6117659
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melaine C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 21
LENGTH: 5970
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-320-878-21

Query Match 4.3%; Score 76.2; DB 3; Length 5970;
Best Local Similarity 44.7%; Pred. No. 2,2e-06;
Matches 334; Conservative 1; Mismatches 409; Indels 3; Gaps 1;

QY 32 CCATGTCCTCCATGCGATTCGACGCTTCTGACCTGCTGCTGCGGGGCGACACCTGATCG 91
DB 781 CCACTGCAACGCCACGCGCGGCTCCAGCTCTCGCGCACGCGCGGCTTCACGCGG 722
QY 92 ACGGAGCAACACCGCGGGGGGGCGGCGGACCTGGGCGCTGGCGGCGACCGATCGCG 151
DB 721 AAGTATATATGCGCTGATGACGCTTGGCGCGCACCCCGACGCACTGCGCTGATCGGCC 662
QY 152 CCATGCGGATGCTGCGGACGCGCGCGACACCGCGGCTGACGCTGCGGGCGTGGG 211
DB 661 TCACCGCGGCTGCTGCGCGACGACCGGACACCGGCACTCGACCGGACGAGTGG 602
QY 212 TCGGCGCGGCTTCATGACGCTGACACCGACGACGACAACTGCTCAGGCGTGGCG 271
DB 601 CCGCGCGGCTACACCGCGACCTCGGCGGTGCTGCGGCTTCACCTGCGGGCGCGCT 542
QY 272 ACATACCGCCCAAGATTCGCGAGGCGGCTACACGCTGTGACGGGCAATTGGGCAATCA 331
DB 541 GCGCGCGGACGACGCTGCGGAGGTGCGCGAGACACGCGCTGCGGCTGTACTTCGAGC 482
QY 332 GCGTGGCGCGCTGCGGACGCGCACCGCGCGCGCGCGCTGACCTGAGCTGAGGAGGAG 391
DB 481 CCGGCGACGCGCTGCGGCTGCGGCGGCTGACGCGCGCGCGCGCGGACCTCGCGGAGCG 422
QY 392 GCTTACCGTTGAGGCGCTTGGCGGCTGACCTGAGACGCTTGGCGGCGACGCGCGCG 451
DB 421 AGGTCTTCAGCTTC---CACGCGCACCAAGGCGCTGACGCTTTCGAGGCGGCGCGTGG 365
QY 452 CCGTAAAGCGCGCGCTGTATGCTGGGCGCATTCACGCTGCGCGCGCGGCTGATCGCGACT 511
DB 364 TCACCGAGAGCGCGCGACGCTGCGCGCGCGGATCGCGCGCTTCACCAACTTCCTTGACC 305
QY 512 TGCACGCGCGCGCGCACGCGAGAAATGCGGCGCATGCGGAGACCTGCGGCGAGAACCA 571

DB 304 TGCCCGCGCGCACCGCCCGCGCGGAGACCAACGCAAGATGACGAGCGCGCGCGCA 245
QY 572 TGCGCAGCGCGCGCAATCGGATTCGACCGGCGCTTCTACCGCGCGCGCGCGCGCA 631
DB 244 TGCGCCTACCTCTCTCGAGCGCTTCCGAGTCTATCGACCGGAACCGCGCGCAACG 185
QY 632 CCACCGAAGATATCGAGGCTGCGCGCGCTGAGCGCGGACGCGGCGCATCTACGCA 691
DB 184 CCGSCTACCGCGAGCACCTCGCGGACCTCGCGCGCTCTCGTCCGCGGACACGAGCGCG 125
QY 692 CCACATGCGCGCGACGAGGCGGACGACATGCTGCGCGCGCTGAGGAGAACTTCGATCG 751
DB 124 ACGGCTCAACCAACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 65
QY 752 GCGCGAGCTGAGCTGCGGCTGATG 778
DB 64 ACGGACGCTGCTGATGAGCTCTGA 38

RESULT 6
US-08-804-227C-1
Sequence 1, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuntz, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS

QY	35	TCGCCAAATCCGATTCCCAAGCCCTTGACACTGCTGTGCGGGGGCCACCTTCATGAGC	94
Db	28810	TGCCCCGGCCGGCCCGCCGAGAGACTCTTACCGCCCGCCGACACACGCTTGC	28865
QY	95	GCACGCAACACCCCGGGGCGCGCGCCAGCCTTGCGCTGCGGCGCACCGCATGCGCCGA	154
Db	28870	GCGTCAACCCCGCGGGGGGAGACGCTGGAGGAATGTCGTCTGGGCCGCCCGGACCGG	289229
QY	155	TCGGGATCTGTGTGAGACGCCCGCCGCGCAACCCGGGTCCAGTGTGCGGCTGTGTGTG	214
Db	28930	AGGAACTATGTCGACCGGGGCAAGTAAAGCATGCGCTCCGCGCGGCGGCGTGAATCTCC	28989
QY	215	CGCCCGGCTTCATCGACTGCGACACCCAGACAGACAAATCACTGCTCAGGCTCCGACAA	274
Db	28990	GGGAGCCCTGATCGCCCTTCGCGCATGTATCCCGGGCAA-----GGAAACATAGGGCGCG	29043
QY	275	TGACGCCCAAGATCTTCGAGGGCGTCACACAGGTGTCTACGCGGCAATTGCGCATATGCGC	334
Db	29044	AGGGAACCCGCGCTGTCGTCTGAGAACCGGCCCGGATGTACACGCGCTCTCCGCGGAGAC	291030
QY	335	TGGCGCGGTGGCGACAGCCCAACCCGCCCGCCCTTGACCTGCTGTGACGAGGCGGT	394
Db	29104	GCGTGTCTGCGCATGTGGAAACGGCGGCTTTCGGGCCCTCTGTGTGTGCGCACACCATATG	29163
QY	395	CTTACCGTTTCAGCGCTTCGCGCATCTGCGCATCTGGACGCGTGTGGGCGCAAGCGCGGGCG	454
Db	29164	TGGCCCGCATCCCGACGCGCTGTGTGTGTAAGCGGAGCGGCTCTCGTCCCGCGGTCTCC	29223
QY	455	TCAAGCGCGCTCTGTATGTGTGGCCATTCAAGCGTGGCGCGCGCGGTCTATGCGCGACTTGC	514
Db	29224	TCACCTCTTACTACACCGCTGTACCCCGGCTGTGGCCCGGGCCGCGACCGAGAACGTCCTCG	29283
QY	515	AGCGGCGCGCACCGACGAGGAAATCGGGGCGCATCGGGAGACTGGCCGAGAACCATATG	574
Db	29284	TCCACGCCCGCGCGCGCGCTGTGCGCATGGGAGACCTTCAACTCCCGCCACACTCGGCGC	29343
QY	575	CCAGCGGCGCCATCGCATTTTCAGCCGCGGCTTTCACCGCGCGCGCGCGCGCGCGCAACA	634
Db	29344	TGGAGGTGTACGCCACCGCGAGACACCGGCAATGGAGACGCCCTGCAGAGAACGCGCATCC	29403
QY	635	CCGAAAGATATTCAGAGGTGTGCGCGCGCGCTGAGACGCGCATGGGCGCATCTACGCCAACCC	694
Db	29404	CCGAGACCGGATCGCCGACACTCCGCAACCTG---GACTTGGCGAGGCGTTCCTGTGCC	29460
QY	695	ACATGCGGAGAGAGGGGAGACATCTGTGGCGCGCGCTGGAGAAACCTTCGCAATCGGCC	754
Db	29461	GGACGGGCGCGCGGGGTGTGACATCTGTGTGAACTCTTCGGCGCGGAGATTCGTGACG	29520
QY	755	GCGAGCTGGAGAGTCCCGGTGTGTATCTCGCACACCAAGATCATGGGCGCAAGCCCAATTTCG	814
Db	29521	CCTCAGCTGCGGCTCTGTCGCGCGCGCGCGGCACTTCTTGGAACCTGGGAAAGCCGCACTCC	29580
QY	815	GCCGCTGCGCGAGACGCTGCGGCTGTATCGAGAGCGCGCATGGCGCGCCAGACAGCTGCGC	874
Db	29581	GCGACCCCGCGCGGATGCGCGCGCGCCATCCGGGCAACGACTACCGGGCGTTTGACACTGG	29640
QY	875	TGGACGGGTATTCCTTACTGTGGCGCGGCTTCACACATGCTCAAGACAGAGACGGGTGCTGTG	934
Db	29641	TGCAGGCGGGTCCCGACACCGGTGGGAGATGTCTCGGGAACTGTGGAATCTGTTCGCGG	29700
QY	935	CCGAGCCACCATATCATCTGATGAGGAAGGCGCTTCCCGGAATGATGAGGGGCGGACCTGG	994
Db	29701	CCGAGGGGCTGTGCGCGCGCTGCGCTACCGGCTTACGGCATATGCGAGACCGCGCACCC---	29756
QY	995	ATGAAGTCCGGCGGAGCGCGGCAATTCACAGTACGACTGTGTGCCGAGCTGTGAGCGCGG	1054

Db	29757	----	GCCTTGGGACACCCCTCACCACAGGCCCGGACACCCGGCAAGCTGTGCTGCAGCGTCC	29811
QY	1055	CCGGGGCCACTTACTTTCATGATGAGCAACCAGACCTGCAGGCGATCTGGCGTTGGGCG	1114	
Db	29812	CTGGCGGATTTGCACACCCACCCAGCGTGTCTCTCTACACGGCGGACAGCGACAGCTDGGCC	29871	
QY	1115	GCACCATGATTCGGCTTCGCCAGCGCTTCGCCACAGCAGCGCCCGCATTCGCGCCTGTGGG	1174	
Db	29872	ACACACTCGCCCGCCCATCTGGTGTCAACCCGACAGCGGGGTACGGCACACCTGCTGTGCGCGGCC	29931	
QY	1175	GCACATTCGCGGGGTGGTGGGAGCACTATAGCGCGGACCTGGGGCCTGTTCGCCGTGAGAA	1234	
Db	29932	GCACCGGCGCGCGCGGCCCGACAGGCGCGTCGCGGAATGATATGCTAACTTGGGCGAGTTGGCG	29991	
QY	1235	CGGCGGTATGGAAGATGACCCGGCCTTGACCGCGCGCGCTTGGAGCTGGCGG--GGCGCG	1291	
Db	29992	CCGAGGTCCGGGTTCGCGGCGCTCGACAGCGGCGCGACCGGACGGCGGTCAACCGAACTCCTCG	30051	
QY	1292	GGCAGCTGCAGGCCCGGGTACTTGGCCGACCTGGTGGTGTTCGACCCGGGACAGGTGGCGG	1351	
Db	30052	CCGGAAATCCCGTCAGACACCCGCTCGGCGCGCGTGTCTCACGCGCGGGACACTCGAAG	30111	
QY	1352	ATACCGGACACCTTCGAACACCCCTACCGAGCGCGCGCGGATTCATTCCTCGTGTACGTCA	1411	
Db	30112	ACGGCACCATTCCTCTCACTGTACACCGCGGAAACATGCAACAGTGTGGCGCCCAAGGCGG	30171	
QY	1412	ACGGCGCGCGGATCTGGCAGAGCAGCGGGTTCACCGCGCAGCATCCGCGCGCGTGTGG	1471	
Db	30172	ACGGCGTGTCTCACTGTGACAGAGGTGACCCCGGAGCGCGACCTCTCGGCGTTCTGTCTC-	30230	
QY	1472	CACGCAAGGCGCGCTGAGACCCGCGCGCACGCCCTTACAAATCGGCGGTGAAGGGGCGCGGT	1531	
Db	30231	--TACTGTCTCTCTCGCGCGTGTGTGGCACCCCGGCCAAGGCGGCTTACGCCGCGGGCCA	30288	
QY	1532	GCGCGCCCTCTCCAAACCTTGAAGCAAAAGCGTCAATGAGGCCCTTCCCTCCCTCGCAATA	1591	
Db	30289	ACGGCTTCTCTGGAGCGGCTTCGCCCGATACCGCAAGGCGCTTGGGCTGTGCGCGCTTCTGC	30348	
QY	1592	CGGCGCCACCCGATATGTTGGGCAAGAAAGTATGGGCGCGCGCTGTGGCGCCACAGGCA	1651	
Db	30349	TGGCTGTGGGACTGTGGGGGACGAAACAGCCGCAATGGGGGGGCACTCGAACATGTGGGCA	30408	
QY	1652	AGGCGCGGAAATATACCTCTGCAACATCTGTGCAAGGCCAGCGGCAATGCGGCTTCGACCC	1711	
Db	30409	TGCACAGGCGCGCTGTAACCGGAGCGGCATCATGATGCGCTCACGACGCGAGGGGCTTCGCC	30468	
QY	1712	TGTCCAGGCGCGAGCTGGGC	1731	
Db	30469	TGTTTCAGCGCGCACAGGAC	30488	

RESULT 7
 US-08-461-775-8
 : Sequence 8, Application US/08461775.
 : Patent No. 5838773
 :
 : GENERAL INFORMATION:
 : APPLICANT: MAZODIER, Philippe
 : APPLICANT: GUGLIEMI, Gerard
 : TITLE OF INVENTION: REGULATORY NOCLECTIVE SEQUENCE OF THE
 : TITLE OF INVENTION: INITIATION OF TRANSCRIPTION
 : NUMBER OF SEQUENCES: 15
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Burns, Doane, Swecker & Mathis
 : STREET: George Mason Bldg., Washington & Prince Sts.
 : CITY: Alexandria
 : STATE: Virginia
 : COUNTRY: United States
 : ZIP: 22313-1404
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,775
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/050,313
FILING DATE: 10-MAY-1993
APPLICATION NUMBER: FR 9011186
FILING DATE: 10-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 010830-035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-2021
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1320
US-08-461-775-8

Query Match      4.38; Score 75; DB 2; Length 1320;
Best Local Similarity 46.9%; Pred. No. 3.2e-06;
Matches 234; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 297 CGTCCACGAGGTGTGACAGGCGCAATGCGGCATGAGCCTGGCCGCCGCTGGCGGACGCCAA 356
DB 90 CGGCCCCAAGAGGCGCGCAACGCTGATCGACAGAAAGTTGCGGCCCGCCGACCATCACCAG 149
QY 357 CCGGCCCCCGCTGTGACCTGCTGGAAGAGGCGCTCTTACCGCTTTCGAGCGCTTTCG 416
DB 150 CGAGCGGCTGACATCGCCCGTGTGAGTGTGAGTGAGAGAGCCGATGACAGAAAGTCTGGCGC 209
QY 417 CGACTACCTGAGGCGCTGTGGGGGCGGCGGCGGCGGCGCTCAAGCGCCGCTGTATGTGG 476
DB 210 CGAGCTGTGTAAGAGGTGTGCGAGCAAGACCAAGACATGCGGGGTAGCGGACCAAC 269
QY 477 CCATTCAACGCTGCGGCGCGGCTGATGCGGAGCTTGCACGCGCGCCCGCCACCGAGCA 536
DB 270 CGGAGCCGTGCTGCGGCGGCGGCGGCTGCTGCGGAGGCGCTGCGCAACGTCGCGCGCGCGC 329
QY 537 AATCGCGGCCATCGGAGACCTGCGCGGAGAGGACCATGCGGCGGCGGCGGCGGCGGCGG 596
DB 330 CTCGCCGCGCGCGCTGAGAAAGGCGATGAGCGCGCGCGCTGCGCGCGCTGCGCGGAGCT 389
QY 597 GACCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 656
DB 390 GCTCGACACCGCGCGCGCTGAGCAAGTCCGACATGCGCGCGCTGCGCGCGCTGCGCGGCTC 449
QY 657 CCGGCGCGCTGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 716
DB 450 CGGCGGAGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 509
QY 717 CATCGTGGCGCGGCTGAGAGAACTTCCGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 776
DB 510 CGGTGTGATCATCGCTGAGAGAGTCCAAACGCTTGGGTGCGACTGGAATTTCACCGAGGG 569
QY 777 GATTCGAGCAACCAAGGTC 795
DB 570 CATGGCTTTCGACAAAGGCG 588

```

RESULT 8
 US-09-031-606-8
 ; Sequence 8, Application US/09031606
 ; Patent No. 6153404

```

GENERAL INFORMATION:
APPLICANT: MAZODIER, Philippe
APPLICANT: GUGLIENT, Gerard
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
TITLE OF INVENTION: INITIATION OF TRANSCRIPTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: George Mason Bldg., Washington & Prince Sts.
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,606
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/050,313
FILING DATE: 10-MAY-1993
APPLICATION NUMBER: FR 9011186
FILING DATE: 10-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 010830-035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-2021
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1320
US-09-031-606-8

Query Match      4.38; Score 75; DB 3; Length 1320;
Best Local Similarity 46.9%; Pred. No. 3.2e-06;
Matches 234; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 297 CGTCCACGAGGTGTGACAGGCGCAATGCGGCATGAGCCTGGCCGCCGCTGGCGGACGCCAA 356
DB 90 CGGCCCCAAGAGGCGCGCAACGCTGATCGACAGAAAGTTGCGGCCCGCCGACCATCACCAG 149
QY 357 CCGGCCCCCGCTGTGACCTGCTGGAAGAGGCGCTCTTACCGCTTTCGAGCGCTTTCG 416
DB 150 CGAGCGGCTGACATCGCCCGTGTGAGTGTGAGTGAGAGAGCCGATGACAGAAAGTCTGGCGC 209
QY 417 CGACTACCTGAGGCGCTGTGGGGGCGGCGGCGGCGGCGGCTCAAGCGCCGCTGTATGTGG 476
DB 210 CGAGCTGTGTAAGAGGTGTGCGAGCAAGACCAAGACATGCGGGGTAGCGGACCAAC 269
QY 477 CCATTCAACGCTGCGGCGCGGCTGATGCGGAGCTTGCACGCGCGCCCGCCACCGAGCA 536
DB 270 CGGAGCCGTGCTGCGGCGGCGGCGGCTGCTGCGGAGGCGCTGCGCAACGTCGCGCGCGCGC 329
QY 537 AATCGCGGCCATCGGAGACCTGCGCGGAGAGGACCATGCGGCGGCGGCGGCGGCGGCGG 596
DB 330 CTCGCCGCGCGCGCTGAGAAAGGCGATGAGCGCGCGCGCTGCGCGCGCTGCGCGGAGCT 389
QY 597 GACCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 656
DB 390 GCTCGACACCGCGCGCGCTGAGCAAGTCCGACATGCGCGCGCTGCGCGCGCTGCGCGGCTC 449

```


QY 657 CCGGCCGCTGAGGCGCATGGGCGCATTCACGCCACACCATGCGGAGCAAGGCGAGCA 716
DB 450 CCGCGAGGACAAAGCAGGTGGCGAGCTCATGCCGAGGCGAAGGAGCAAGGTGGCGAAGCA 509
QY 717 CATCGTGGCGGCGCTGGAGAAACCTTCCGATCGGCGCGGAGCTGGAGCTGGCGGTGGT 776
DB 510 CGGTGTCATCACCGTCGAGAGAGTCCACACACCTTCGCTGCACTGACCTGACCGAGGG 569
QY 777 GATCTCGCACCAAGATC 795
DB 570 CATGGCCTTCGACAAAGGCG 588

RESULT 9
US-08-461-775-10
Sequence 10, Application US/08461775
Patent No. 5858773
GENERAL INFORMATION:
APPLICANT: MAZODIER, Philippe
APPLICANT: GUGLIERI, Gerard
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: George Mason Bldg., Washington & Prince Sts.
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461.775
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/050.313
FILING DATE: 10-MAY-1993
APPLICATION NUMBER: FR 9011186
FILING DATE: 10-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 010830-035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1620 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1620
US-08-461-775-10

Query Match 4.3%; Score 75; DB 2; Length 1620;
Best Local Similarity 46.9%; Pred. No. 3.3e-06;
Matches 234; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 297 CGTCACACGCGTGTGTCAGGCGCAATGGCGCATCAGCGCGCGCGTGGCGGCGAGCGCA 356
DB 90 CCGGCCGAGAGGCGGCGCAAGTCTCATCGACAAAGAGTTGGGCGCGCGCGCATCGACAA 149
QY 357 CCGCGCGCGCGCGCGCTGCTGTCGACGAGAGCGGCTTACCGCTTTCAGCGCTTCGC 416

DB 150 CCAGCGGCTTCACCATCGCCCGTAGAGTGCAGAGCAGCCGTAACAGAACCTCGGCC 209
QY 417 CGACTACCTGAGACGGGTTCCGGGCGACAGCGCGCGCGCTTCAACGCCCTGTATGGTGG 476
DB 210 CCAGCTGTCGAAGAGGTGGCGCAAGACCAACGACATCGGGGTGACGGGCGACACAC 269
QY 477 CCATTCACAGCTGCGCGCGCGCGGTGATGCGGAGCTTGAGCGCGCGCGCGCGAGAGA 536
DB 270 CGCGACCGTGTGGGCG 329
QY 537 AATCGCGGCATGCGGAGCTTGGCGGAGAGCCATGCGCGCGCGCGCGCGCGCGCGCG 596
DB 330 CTCGCCGAG 389
QY 597 GACCGCGCGCTTCACCG 656
DB 390 GCTCGACACCGCTC 449
QY 657 CCGGCCGCTGAGCGCGCATGGCGGCGATCTACGCCACACCATGCGCGCGAGAGCGAGCA 716
DB 450 CCGCGAGGACAAAGCAGGTGGCGAGCTCATGCGGAGGCGAGTGCAGCAAGGTGGCGA 509
QY 717 CATGTGCGCGCGCTGGAGAAACCTTCGCGATCGCGCGCGCGCGCGCGCGCGCGTGT 776
DB 510 CGGTGTCATCACCGTCGAGAGTCCACACACCTTCGCTGCACTGACCTGACCGAGGG 569
QY 777 GATCTCGCACCAAGATC 795
DB 570 CATGGCCTTCGACAAAGGCG 588

RESULT 10
US-09-031-606-10
Sequence 10, Application US/09031606
Patent No. 6151404
GENERAL INFORMATION:
APPLICANT: MAZODIER, Philippe
APPLICANT: GUGLIERI, Gerard
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: George Mason Bldg., Washington & Prince Sts.
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031.606
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/050.313
FILING DATE: 10-MAY-1993
APPLICATION NUMBER: FR 9011186
FILING DATE: 10-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 010830-035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1620 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1620
US-09-031-606-10

Query Match 4.38; Score 75; DB 3; Length 1620;
Best Local Similarity 46.98; Pred. No. 3.3e-06;
Matches 234; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY CGTACACAGGCTGATGACGCGGCAATGCGGATGAGCTGCGCGCGCTGGCGGACGCCAA 356
DB CGGCCCCAGGCGCGGACGCTGCTATGACGAGAGTTCCGCGCGCGCGGACCATCACC 149
QY CCGCGCGCGCGCGCGCTGACCTGCTGACGAGGCGGCTTACCGCTTTCAGCGCGTTCG 416
DB CGAGCGGCTACCATGCGCGCGCTGAGTGCAGTGCAGACCGCGTACGAGAACTTCGCG 209
QY CGACTACCTGAGCGCTGCGCGCGCGCGCGCGCGCGCTGATGCTGAGTGG 476
DB CCACTGCTGACGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 269
QY CCAATCAAGCTGCG 329
DB CGGAGCGCTGCTGCG 399
QY AATCG 596
DB CCG 389
QY GACCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656
DB GCTCGACACCG 449
QY CCG 716
DB CGGCGCGAGACGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 509
QY CATCGTGGCG 776
DB CGGTGCTATCATCGCTGAGAGAGTCAACACTTGTGTGACTGATCTTCAACCGAGG 569
QY GATCTGCGCACCAAGGTC 795
DB 570 CATGGCTTCGACAGGCG 588

RESULT 11

US-08-458-568A-11/C
Sequence 11, Application US/08458568A
Patent No. 5821339
GENERAL INFORMATION:
APPLICANT: Schaffer, Priscilla A.
APPLICANT: Yeh, Lily
TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/458,568A
FILING DATE: 02-JUNE-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/065,146
FILING DATE: 05-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Leary Ph.D., Kathryn R.
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: DECT-0029
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 12001 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Herpes simplex virus
STRAIN: Herpes Simplex Virus Type 1
US-08-458-568A-11

Query Match 4.28; Score 74.4; DB 1; Length 12001;
Best Local Similarity 42.65; Pred. No. 5.2e-06;
Matches 581; Conservative 0; Mismatches 766; Indels 17; Gaps 3;

QY 189 GGTGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 248
DB 5125 GGTGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5066
QY 249 CAACTACCTGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 308
DB 5065 CAGCAGCTCTTCAATCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5006
QY 309 GGTGACG 368
DB 5005 GGTGACG 4946
QY 369 CTTGACCTGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 428
DB 4945 CCAATACG 4886
QY 429 CCGCTGCG 488
DB 4885 CCGTGTGCG 4826
QY 489 GCG 548
DB 4825 AGATGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4766
QY 549 GCGGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 603
DB 4765 CCGGTACG 4706
QY 604 GCGTTCACCG 663
DB 4705 CCG 4646
QY 664 CCGGCG 723
DB 4645 CCG 4586
QY 724 GCG 783
DB 4585 GAGCG 4526
QY 784 CACCAAGGTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 843

Db 5810 GAGGCCCTACGAGAGCCCGGCTTGAACCCGCGGCGGACCCGGCTGCTGAGG 5869
QY 1249 ATGACGGGCTGACCGCGCGGCTTGGC-----CTGGCGGCGCGGCGAGCTGCA 1302
Db 5870 ACCAGGTCGAGGTGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5929
QY 1303 GCCGGGTACT--TCGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1359
Db 5930 GACGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5989
QY 1350 ACCCTGCAACACCTTACGAGCGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 1419
Db 5990 TTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6049
QY 1420 CCGGCTGCGAGAGAGAGCGCTTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1479
Db 6050 ACCCTGAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6109
QY 1480 GCCGCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1516
Db 6110 AGCGCGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6146

RESULT 13

US-08-461-775-9
; Sequence 9, Application US/08461775
; Patent No. 5858773
; GENERAL INFORMATION:
; APPLICANT: MAZODIER, Philippe
; APPLICANT: GUGLIEMI, Gerard
; TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,775
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/050,313
; FILING DATE: 10-MAY-1993
; APPLICATION NUMBER: FR 9011186
; FILING DATE: 10-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2167 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-461-775-9

Query Match 4.2%; Score 73.4; DB 2; Length 2167;
Best Local Similarity 46.7%; Pred. No. 6.8e-06;

Matches 233; Conservative 0; Mismatches 266; Indels 0; Gaps 0;
QY 297 CGTACACAGGTGCTGACGCGGCAATTCGCGCATCAGCTGGCGCGCTGCGCGACGCCAA 356
Db 938 CGGCCCCAAGGGCGCGCAAGTGTGATGACAAAGATTGCGGGGCCGACCATCAACGA 997
QY 357 CCGGCCCCCGCTGAGCTGCTGAGCAAGAGCGGCTTTACCGTTTGAGCGCTTGC 416
Db 998 CGAGGCGGTACCATGCGCGCTGAGAGGTGAGTGCAGTGCAGACCGCTACAGAACCTCGCGCG 1057
QY 417 CGACTACCTGAGAGCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
Db 1058 CCAAGCTGTCAAGAGAGGTGCGACCAAGACCAAGACATCGGGGTGAGGAGCAACAC 1117
QY 477 CCATTCAAGCTGCG 536
Db 1118 CGGAGCGGTGCG 1177
QY 537 AATCGGCGCATGCGGCG 596
Db 1178 CTCCCG 1237
QY 597 GACCGCGCGCTTCTACCG 656
Db 1238 GCTCGACACCG 1297
QY 657 CCGGCGCGTGAAGCG 716
Db 1298 CCGGAGAGCAAGCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1357
QY 717 CATCGGCG 776
Db 1358 CGGTGATCAACAGTGCAGAGAGTCCACACCTTGGTGTGACCTGACTTCCAGGAGG 1417
QY 777 GATCTGCACACCAAGTGC 795
Db 1418 CATGCGCTTGCACAAAGGC 1436

RESULT 14

US-09-031-606-9
; Sequence 9, Application US/09031606
; Patent No. 6153404
; GENERAL INFORMATION:
; APPLICANT: MAZODIER, Philippe
; APPLICANT: GUGLIEMI, Gerard
; TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,606
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/050,313
; FILING DATE: 10-MAY-1993
; APPLICATION NUMBER: FR 9011186
; FILING DATE: 10-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113

REFERENCE/DOCKET NUMBER: 010830-035
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2167 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-09-031-606-9

Query Match 4.2%; Score 73.4; DB 3; Length 2167;
 Best Local Similarity 46.7%; Pred. No. 6.8e-06;
 Matches 233; Conservative 0; Mismatches 266; Indels 0; Gaps 0;

297 CGTACACAGGTGTGACAGGCAATTGGCGCATCAGCTGGCGGGCGGGGACGCCAA 356
 Db CGGCCCCAGGCGCGACGTGTCATCGACAGAAAGTTGGCGCCCGGACCATCACC 997
 357 CCGCGCGCGCGCGCGACCTGCTGACAGAAAGCGGCTTTACCGTTTGAAGCGCTTGC 416
 Db CGAGCGGCTCACATCGCCCGTGAAGTGCAGTGCAGACCGCGTACAGAACTCGGCGC 1057
 417 CGACTACCTGACCGCTTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
 Db CCAGCTGTCAGAGAGGTGGCGACCAAGACCAAGACATCGCGGCTGACGGCACACAC 1117
 477 CCATTCACGCTGCG 536
 Db CGGACCGCTGCG 1177
 537 AATCGCGCGCATGCGGGAAGCTGCGAGAGAGCATGCGCGCGCGCGCGCGCGCGCG 596
 Db CTCCCG 1237
 597 GACCGCGCGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656
 Db GCTCGACCG 1297
 657 CCGCGCGGTGAGCG 716
 Db CGCGCGAGCAAGAGGTGGGAGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1357
 717 CATCGTGGCG 776
 Db CGGTGTATCATACGTGAGAGATCCAAACCTTGTGCTGACCTGACTGACGAGGG 1417
 777 GATTCGACCAACAAGTTC 795
 Db 1418 CATGCGCTTCGACAAGGC 1436

RESULT 15
 US-08-461-775-11
 Sequence 11, Application US/08461775
 Patent No. 5858773
 GENERAL INFORMATION:
 APPLICANT: MAZODIER, Philippe
 APPLICANT: GUGLIERI, Gerard
 TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Burns, Doane, Swecker & Mathis
 STREET: George Mason Bldg., Washington & Prince Sts.
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/461,775
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/050,313
 FILING DATE: 10-MAY-1993
 APPLICATION NUMBER: FR 9011186
 FILING DATE: 10-SEP-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Crane Feury, Sharon E
 REGISTRATION NUMBER: 36,113
 REFERENCE/DOCKET NUMBER: 010830-035
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2668 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-461-775-11

Query Match 4.2%; Score 73.4; DB 2; Length 2668;
 Best Local Similarity 46.7%; Pred. No. 7e-06;
 Matches 233; Conservative 0; Mismatches 266; Indels 0; Gaps 0;

297 CGTACACAGGTGTGACAGGCAATTGGCGCATCAGCTGGCGCGCGCGCGCGCGCGCA 356
 Db CGGCCCCAGGCGCGACGTGTCATCGACAGAAAGTTGGCGCCCGGACCATCACC 997
 357 CCGCGCGCGCGCGCGACCTGCTGACAGAAAGCGGCTTTACCGTTTGAAGCGCTTGC 416
 Db CGAGCGGCTCACATCGCCCGTGAAGTGCAGTGCAGACCGCGTACAGAACTCGGCGC 1057
 417 CGACTACCTGACCGCTTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
 Db CCAGCTGTCAGAGAGGTGGCGACCAAGACCAAGACATCGCGGCTGACGGCACAC 1117
 477 CCATTCACGCTGCG 536
 Db CGGACCGCTGCG 1177
 537 AATCGCGCGCATGCGGGAAGCTGCGAGAGAGCATGCGCGCGCGCGCGCGCGCGCG 596
 Db CTCCCG 1237
 597 GACCGCGCGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656
 Db GCTCGACCG 1297
 657 CCGCGCGGTGAGCG 716
 Db CGCGCGAGCAAGAGGTGGGAGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1357
 717 CATCGTGGCG 776
 Db CGGTGTATCATACGTGAGAGATCCAAACCTTGTGCTGACCTGACTGACGAGGG 1417
 777 GATTCGACCAACAAGTTC 795
 Db 1418 CATGCGCTTCGACAAGGC 1436

Search completed: May 11, 2003, 10:52:38
 Job time: 235 secs


```

OTHER INFORMATION: ORF 7; positive strandedness
NAME/KEY: misc_feature
LOCATION: (12751)..(10829)
OTHER INFORMATION: ORF 8; negative strandedness
NAME/KEY: misc_feature
LOCATION: (13617)..(12802)
OTHER INFORMATION: ORF 9; negative strandedness
NAME/KEY: misc_feature
LOCATION: (15203)..(13614)
OTHER INFORMATION: ORF 10; negative strandedness
NAME/KEY: misc_feature
LOCATION: (15591)..(15863)
OTHER INFORMATION: ORF 11; positive strandedness
NAME/KEY: misc_feature
LOCATION: (15880)..(19035)
OTHER INFORMATION: ORF 12; positive strandedness
NAME/KEY: misc_feature
LOCATION: (19032)..(39713)
OTHER INFORMATION: ORF 13; positive strandedness
NAME/KEY: misc_feature
LOCATION: (39713)..(65800)
OTHER INFORMATION: ORF 14; positive strandedness
NAME/KEY: misc_feature
LOCATION: (65826)..(66530)
OTHER INFORMATION: ORF 15; positive strandedness
NAME/KEY: misc_feature
LOCATION: (66546)..(67370)
OTHER INFORMATION: ORF 16; positive strandedness
NAME/KEY: misc_feature
LOCATION: (67384)..(70059)
OTHER INFORMATION: ORF 17; positive strandedness
NAME/KEY: misc_feature
LOCATION: (70099)..(70662)
OTHER INFORMATION: ORF 18; positive strandedness
NAME/KEY: misc_feature
LOCATION: (70659)..(71906)
OTHER INFORMATION: ORF 19; positive strandedness
NAME/KEY: misc_feature
LOCATION: (73439)..(71964)
OTHER INFORMATION: ORF 20; negative strandedness
NAME/KEY: misc_feature
LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
NAME/KEY: misc_feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
NAME/KEY: misc_feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
NAME/KEY: misc_feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
NAME/KEY: misc_feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
NAME/KEY: misc_feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
NAME/KEY: misc_feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
NAME/KEY: misc_feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
NAME/KEY: misc_feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
NAME/KEY: misc_feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness

```

```

NAME/KEY: misc_feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; positive strandedness
NAME/KEY: misc_feature
LOCATION: (87484)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-976-059-1

```

```

Query Match          4.9% Score 86; DB 9; Length 88421;
Best Local Similarity 43.1%; Pred. No. 1e-10;
Matches 605; Conservative 0; Mismatches 775; Indels 24; Gaps 3;

```

```

QY 35 TGTCCCATCCGATTCCTCCAGCCCTTCGACCTCTGCTCGCGGCGCCACCTCATCGACG 94
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46230 TGGCCCGGACAGAGTGTCTGCACAGGTCTTCACAGGTCTCCGACGAGCCGTGACAC 46289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 95 GAGCAACACCCCGGGGGGGGGGGGGGACCTGCGGGGTGCGGGGCGGACCCCATCCGCCA 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46290 ACATCTCGACCCCGGGGGGGGGGGGAGCCCGCGTGGCGGTGACGTGCCGGCGGCC 46349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 155 TCGGCGATCTGTGCGAGCGCGCGCGGACACACCCGGGTGACGTGTCGGGCTGTGTGTCG 214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46350 GGGTCGAGAGAGCGGTCTGCGCGAAGCGCGCGCTTCACCTTCGCGCGCGAGATCC 46409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 215 CGCCCGCTTCATCGACTGCGACACCGACGACGACACTACTGCTCAG-----GCGTC 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46410 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 46469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 269 GGGACATGAGCGCCAGATCTGCGAGGGGTGACACAGGTGTGACAGGGGCAATGGGCA 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46470 ACCACATGCGGGCGCGAGCGGTGTGATGCGCGCGCTGCGCGCGACCTGCGGACCGCT 46529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 329 TCAGCTGCGCGCGCTGCGCGACGCGACGCCGCCCGCGCGCTGCGCTGCGAGAG 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46530 ACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 389 GCGGCTTACCTGTTTCCAGCGCTTGCAGACTACTGAGACCGGTTCGCGCGCGCGCG 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46590 ACTAGCGCTTGTGCGAGCTGACCTGTGGTTCGACAGGACCGCGGACCTGTATCT 46649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 449 CGGCGGTCAAGCGCGCTGTATGTGAGGCAATTCAGGTGCGCGCGCGCGGTATGCCG 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46650 CCGGCGAGATCCGCCACTGTGCGCGAGCGCTGCGAGCGCTGCGAGAGCTGTGACTTC 46709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 509 ACTGCGAGCGCGCGCGCGCGACGACGAGAAATGCGCGCATGCGGCACTGCGCGAGAG 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46710 CCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 46769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 569 CCATGGCCAGCGCGCGCGCATTCGACATTCGACCGCGCGCTTATCCCGCGCGCGCGCG 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46770 TCCCGCGCGAGCTGCGACCGAGACCTGCGCGCGGTGCGCGCGCGCGCGCGAGCT 46829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 629 CCACCACGAGAGATCATCGAGGTGTGCGCGCGCGGTGCGCGCATGCGCGCATCTGCG 688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46830 TATGACCTGTGAGCGCGCGCGCGCGCGGTCTGTGCGGCTGCGCGCGCGCGCGAGAG 46889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 689 CCACCACATGCGCGCGAGAGCGAGACATGCTGCGCGCGCGGTGAGAGAACTTCCGCA 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46890 TCCCGATGCGACCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 46949
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 749 TCGGCGCGAGCTGCGACGTGCGCGGTGTGTATCTGCGACCAAGGTATGAGGCGACCGCA 808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46950 TCTTGCTCAACAGCTGTGCTGCGACACCGACCTGCGCGCGCGCGCGGTGACCGAG 47009
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 809 ATTGCGCGCTGCGCGCGAGAGCGTGCCTGATGAGGCGCGCATGCGCGCGCGAGAGC 868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 47010 TGTCTGCGCGAGTGTGCGCGAGCTGTGACCTTCGCGCGCGCGCGCGCGCGCGGTG 47069
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 869 TCTGCTGAGCGGTATCTTACGTGCGCGGTGCGACCAATGCTCAAGAGAGACCGCGTGC 928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 47070 AAGAGCTGTGAGAGAGCTCAACCGCGCGCGCTC-----GCTGCGCGCGCGCGCGCT 47123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 929 TGTGCGCGAGCGACCATCATCATCTGTGTGCAAGCCCTTCCCGCAACTGAGCGCGCG 988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


? PRIOR FILLING DATE: 1999-06-25
 ? PRIOR APPLICATION NUMBER: US 09/105,537
 ? PRIOR FILING DATE: 1998-06-26
 ? NUMBER OF SEQ ID NOS: 53
 ? SEQ ID NO 7
 ? LENGTH: 1248
 ? TYPE: DNA
 ? ORGANISM: *Streptomyces venezuelae*
 US-09-308-384B-7

Query Match	4.78;	Score 81.8;	DB 9;	Length 1248;
Best Local Similarity	44.28;	Pred. No. 1.4e-09;		
Matches 439;	Conservative	0;	Mismatches 542;	Indels 12

QY	32	CGATGTCCTCAATCCGATTCCGACGCCCTTCGACTGCTGCTCGCGGGCGGCACCCATCG	91
Db	215	CCACCTGCAACGCTCAACGGCGGGGCTTCAGACTCTTCGGCCGACGGCCGCGCTCAACGGCG	27
QY	92	ACGGCAGCAACACCCCGGGGGCGGGCGGCAGACTCTGGGGGTGCGGGCGCACCGCATCGCG	157
Db	275	AGTGATCATGGCGGTGATGAGAGCTTCGCGGCCACCCCGCACGCACTGGCTGGATCGGCC	334
QY	152	CCATGGCGATTTGTTCGAGCGCGCGGCGCACACCCGGGTGAGAGTGTGGGCGCTGGGG	211
Db	335	TCACCCCGGCTTCCTCCGACATCGACCCGGAGACCGGCACTGAGACCCGGACCAAGGTGG	399
QY	212	TCGGCGCGGGCTTATTCGACTTCGACACCCACGAGACACTTACTCTGCTAGGGCTGCG	277
Db	395	CGCGCGCGGTCAACACCCCGGCACCTCGGGCGTGTCTGGGGGTCACTCTGGGGCGGCCCT	454
QY	272	ACATGACGCCCAAGATCTCGAGAGGGCGTCAACACGAGTGTGTCAGGGGCAATTGGCGCATCA	331
Db	455	GGCGCGCGAGACAGTGGGGAAAGTCCGCCAGACGACAGGCTCGGCTGTACTTTCAGC	514
QY	332	GCCTGGCGCGCTGGCGGACGCCAACCCGCGCGCCCTTGAGACTCTGGAGAACGGCG	391
Db	515	CGCGCAGCGCCCTCGGCTGCGCGCGGTGCGAGCGCGCGCGCGGAGGCTGGGGAGCCCG	574
QY	392	GCTTTACCGTTTGGAGGGCTTCGCCCACTACTCTGAGAGGGTTGGGGCCACGCGCGCG	451
Db	575	AGGTCTTACGTTCA---GCCCAACCAAGGCCGTCAACGCTTCGAGAGGGGGCGCGTGC	631
QY	452	CCGTAAACGCGCGCTGTATGTGTGGGCATTCAACGCTGCGCGCGCGGTATCCGAGCT	511
Db	632	TCACCGAGAGACCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	691
QY	512	TGCAGCGCGCGCCACCGAGAGGAAATCGCGGCCATCGCGGAGCTGGCCGAGGAAGCA	571
Db	692	TGCCCCGGGGCAGCGCCCGCGGGGAGCAACAGCAAGATGAGACGAGGCGCGCGCGCA	751
QY	572	TGGCAGGGGGCGCATTTGACGCGCGCGCTTCAACCCGCGCGCGCGCGCGCGCA	631
Db	752	TGGGCTTCACCTCCCTGAGCGCGTTTCCAGAGTCAATCGACCCGAGCAACCGCACACG	811
QY	632	CCACCGAAGAGATCAATGAGAGGTGTGCCCGCGCGGTGAGCGCGCATGGCGGCAATTCAGCA	691
Db	812	CCGCGTACCGGAGACACTGCGGAGACTCCCGCGCGCTCTCGTGGCGACACAGACCGCG	871
QY	692	CCCACTGCGCGAGAGGCGGACACATCGTGGCGCGCGGTGAGGAAACCTTCGCGATCG	751
Db	872	ACGGGCTCAACAACACACACTAGTGTGTGTGAGATCGACGAGGCGCACACCGGCAATCG	931
QY	752	GCGCGAGCGTGGAGGTGCGCGGTGTATCTCGACACCAAGGTCAATGGGCGAGCCCAAT	811
Db	932	ACCGGACCTCTGATAGAGTCTCTTAAGCGCGAAGCGCTGACACACCCGCGCTACTTCT	991
QY	812	TCGGCGCGCTGCGGAGACGCTGCGCTGATCGAG-----GCCGCAATGGGGCGCG	862
Db	992	CGCGCGGCTCGACAGAGCTGAGACCTTACCGGGGCGACCGCGACGCGCGCGTGGCGGACA	1051
QY	863	AGGACGCTCTGCTGAGCGCGATTCCTACGTGGCGCGGCTCCACATGCTCAAGAGAGCG	922
Db	1052	CCGAGACCCCTCGCGCGCGGCTCTGTCTCTCGTCCGCGACCGGCAACCGCATCGGCGAGCG	1111

QY	923	GCAGCTCTGTCGCGCGGAGACCACTATCTACCTCTGTGTGACGAAGCCCTTCCCGAATGAGCG	982
Db	1112	ACATCTCCCGCGGCTGCCGACCTGCTGTGCTCTGTCCGAGACCGCGCGCCGCGAATGACCG	1171
QY	983	GGCGCGACCTGTGATGATGATGTGCGCGGCGAGCGCG	1015
Db	1172	CGCGCGACCGGAGACGCGGCCCGCGCGGCTCG	1204

RESULT 4
 1015-09-951-780-7

RESULT 4
 US-09-861-289-7
 Sequence 7, Application US/09861289
 Patent No. US20020110897A1
 GENERAL INFORMATION:
 APPLICANT: Sherman, D.H.
 APPLICANT: Liu, H.
 APPLICANT: Xue, Y.
 APPLICANT: Zhao, L.
 TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 FILE REFERENCE: 600,438US1
 CURRENT APPLICATION NUMBER: US/09/861,289
 CURRENT FILING DATE: 2001-05-18
 PRIOR APPLICATION NUMBER: 09/105,537
 PRIOR FILING DATE: 1998-06-26
 NUMBER OF SEQ ID NOS: 43
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 7
 LENGTH: 1248
 TYPE: DNA
 ORGANISM: Streptomyces venezuelae
 US-09-861-289-7

Query Match	4.7%	Score 81.8;	DB 10;	Length 1248;
Best Local Similarity	44.2%;	Pred. No. 1.4e-09;		
Matches 439;	Conservative 0;	Mismatches 542;	Indels 12;	Gaps 2

QY	32	CCATGTCGCAATCCGATTTCCAGGCCCTTCGACTCTCTCTCTGCGGGGCGCACCCCTCATCG	91
Db	215	CCACCTGCGACGCCACGGGCGGGGCTCCAGCTCTCTGCGCAGGCGCGCGGCTTCACACGGCG	27
QY	92	ACGGCAGCAACACCCCGGGGGGCGGCGCGCAACCTGGGGGTGGCGGCGCACCGCATCGCG	15
Db	275	AAGTATCATGCGCGCTGATGAGCTTCCGCCGCCACCCCGCAGCAGCATGCGCTGGATCGGCC	334
QY	152	CCATGGCGATCTGTGGAGCGCGCGCGCGCACACCCGGGTGAGAGTGTGGGCGCTGGTGG	211
Db	335	TCACCCCGGCTTCGCGCCACATCGACCCGGAGACCGGGACACTGACACCCGGACCAAGTGG	39
QY	212	TGCGGCGCGGGCTTACATGACTCTGCGACACCCACGAGACAACTACTCTGCTAGCGGTGCG	27
Db	335	CGGCGCGGGTACACACCCCGCACCTCGGCGCGTGTGTGGGCTTCACCTTGGGGCGGCCCT	45
QY	272	ACATGACGCCCAATATCTCGCAGGGGCGTACACAGGCGTGTACCGGCGAATTGGGGCATCA	33
Db	455	GCGGCGCGGAGACACTCTGCGGAAGTTCGCCACGACACAGGCGTGGCTGTACTTCAAG	514
QY	332	GCTTGGCGCGCTGGCGCAGCCACACCGCGCGCGCCCTCTGAGACTCTGAGACGAGGCG	391
Db	515	CCGGCGACGCGCTCGGCTGGCGGGGTGACAGGGCGGGCGCGGAGCTTGGGGAGCGCG	574
QY	392	GCTCTTACCGTTTGGAGCGCTTCGCCCACTACTGAGACGCGTTGCGGGCCACGCGCGGG	451
Db	575	AGGCTTTCAGGTTTCA---GCGCACCAAGGCGCGTTCACAGGCTTTCGAGGGCGGCGCGTGC	631
QY	452	CCGTCACGCGCGCTGTATGTATGGGGCATTTCAAGCTGAGCGCGCGGGGTATCCCGACT	511
Db	632	TCACCGAGACGGCGACTCGCGCGCGCGGATCCGCGCGCTTCCACAACTTGGCTTGACC	691
QY	512	TGCAAGCGCGCGCACCGGACGAGAGAAATCGCGGCAATGCGGAGCTTGGCGAGAGAAACCA	571
Db	692	TGCGCGGCGCGACCGCGCGCGGGGAGCAACGCAAGATGAGCAGAGGCGCGCGCGCGCA	751

QY 572 TGGCAGGCGGCGCATTCGACATTCGACGCGGCGCTTACACCGCCCGCCGCGCCCA 631
 Db 752 TGGGCTACCTCCCTCCACGCGGTTCCGAGGTATGACCGGAACCGCCCAACACG 811
 QY 632 CCACCGAAGATCATGAGGTGTGCGCGCGCTGAGCGCGCATGCGGCGCATCTACGCA 691
 Db 812 CCGCTTACCGGAGACCTCGCGGACCTCCCGGGCTCTCTGCTCCCGACACGACCGCC 871
 QY 692 CCACATCGGAGACGACGACGACATCTGCGCGCGCTGAGGAAACCTTCGCGATCG 751
 Db 872 ACAGCTTACACACACGACGATGATGATGATGATGATGATGATGATGATGATGATG 931
 QY 752 GCGCGAGCTGAGAGTGGCGGTGATCTGCGACACAGGCTCATGAGGCGGCGGCGCAT 811
 Db 932 ACCGAGACCTGCTATGAGAGTCTGAGAGCGGAGGCGGCGGCGGCGGCGGCGGCG 991
 QY 812 TCGGCGGCTGCGGAGACGCTGCGCGGTGATGAG-----GCCGCAATGCGCGCG 862
 Db 992 CGCGGAGCTGCGGAGAGTGTGAGCGGCTACCGGCGGCGGCGGCGGCGGCGGCGG 1051
 QY 863 AGAGCTGTGCTGAGAGCGGATACCTACGTCGCGCGGCTCCACCATGCTCAAGAGAGC 922
 Db 1052 CCGAGAGCTGCGCGCGCGCGCGCTGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGG 1111
 QY 923 GCGGTGCTGCGCGGAGCGACATCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 982
 Db 1112 ACATCCGCGGAGTGGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1171
 QY 983 GCGGCGACCTGATGAGAGTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1015
 Db 1172 CGGCGCACCGCGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1204

RESULT 5

US-09-988-384B-3/c
 ; Sequence 3, Application US/09988384B
 ; Publication No. US20030073824A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600,536US1
 ; CURRENT APPLICATION NUMBER: US/09/988,384B
 ; CURRENT FILING DATE: 2001-11-19
 ; PRIOR APPLICATION NUMBER: PCT/US99/14398
 ; PRIOR FILING DATE: 1999-06-25
 ; PRIOR APPLICATION NUMBER: US 09/105,537
 ; NUMBER OF SEQ ID NOS: 53
 ; SEQ ID NO 3
 ; LENGTH: 12441
 ; TYPE: DNA
 ; ORGANISM: Streptomyces venezuelae
 ; US-09-988-384B-3

Query Match 4.7%; Score 81.8; DB 9; Length 12441;

Best Local Similarity 44.2%; Pred. No. 1,1e-09;

Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

QY 32 CCATGTCCCATTCGATTCACGACCTTGCAGCTGCTGCTGCGGCGGCGGCGGCGGCGG 91
 Db 11982 CCACCTGACAGCGACGCGCGGCTCCAGCTCTGCGGACGCGCGCGGCTGACCGGCG 11923
 QY 92 ACAGGACGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 151
 Db 11922 AAGTATATATGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 11863
 QY 152 CCATGCGGAGTGTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 211
 Db 11862 TCACCCGCGTCTCCGCGCATGACCGGAGACCGGACGACGACGACGACGACGACGAG 11803

QY 212 TGGCGCGCGGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 271
 Db 11802 CCGCGCGGCTTACACACCGGCGACCTGCGGCGGCTGCGGCGGCTGCGGCGGCGGCGGCG 11743
 QY 272 ACATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 331
 Db 11742 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11683
 QY 332 GCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 391
 Db 11682 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11623
 QY 392 GCTCTTACCTTTCGAGGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 451
 Db 11622 AGGTCTTACGTTTCA---GCCACACAGGCGGCGGCGGCGGCGGCGGCGGCGGCG 11566
 QY 452 CCGTTCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 511
 Db 11565 TCACCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11506
 QY 512 TGCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 571
 Db 11505 TCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11446
 QY 572 TGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631
 Db 11445 TGGCGCTACCTCTCTGAGAGCGGCTTCCAGGTCATGACCGGCGGCGGCGGCGG 11386
 QY 632 CCACCGAAGATCATGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 691
 Db 11385 CCGCTTACCGCGGAGACCTGCGGAGCTCCCGGCGGCGGCGGCGGCGGCGGCGG 11326
 QY 692 CCACATGCGGCGGAGAGGAGGAGACATCTGCGCGCGCGCGCGCGGCGGCGGCGGCG 751
 Db 11325 ACAGGCTTACACACACACAGTACGATGATGATGATGATGATGATGATGATGATG 11266
 QY 752 GCGCGAGCTGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 811
 Db 11265 ACCGCGACCTGTATGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11206
 QY 812 TGGCGCGGCGGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 862
 Db 11205 CCGCGGCGGCGGCGGAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11146
 QY 863 AGAGCTTCTGCTGAGAGCGGATACCTACGTCGCGGCGGCGGCGGCGGCGGCGGCGG 922
 Db 11145 CCGAAGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11086
 QY 923 GCGTGTGCTGCGGCGGAGCGACCATCATCATCTGATGATGATGATGATGATGATG 982
 Db 11085 ACATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 11026
 QY 983 GCGCGACCTGATGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1015
 Db 11025 CGCGCGACCGCGACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 10993

RESULT 6

US-09-860-846-3/c

; Sequence 3, Application US/09860846

; Patent No. US2002016472A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600,438US1

; CURRENT APPLICATION NUMBER: US/09/860,846

; CURRENT FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 3
 LENGTH: 13613
 TYPE: DNA
 ORGANISM: Streptomyces venezuelae
 US-09-860-846-3

Query Match 4.7%; Score 81.8; DB 9; Length 13613;
 Best Local Similarity 44.2%; Pred. No. 1.1e-09;
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

32 CCATGTCCTCCATCCGATTCCTCCAGCCCTTCGACCTGCTCTGCGGGCGGCAACCTCATCG 91
 12669 CCACCTGCAAGCGCACGGCCGGGCTCCAGCTCTCTGCGGACAGCCCGGCTTCACGGCGC 12610
 92 ACGGCAAGCAACACCCCGGGGGGCGCGCGGACCTGGGGTGGCGGCGGACCGCATCGCGC 151
 12609 AAGTATCATGCGCGTGCATGACGTTCGCGCCGACCGCCGACGACACTGCGTGGATCGCGC 12550
 152 CCATCGGGGATCTGTGCGAGCCCGCGCGGACACCCCGGGTGCAGTGTGCGGCGTGTGG 211
 12549 TCACCCGCGGTTCGCGGACATCGACCCGCGACACCGGCAACCTCGAACCCGAGCAGGTGG 12490
 212 TCGGCGCGGCTTCATGACATCGGACACCCGACGACGACACTACTGCTCAGCGGTGCGG 271
 12489 CCGCGGGGTACACACCCCGGACCTCGGCGCTGCTGCGGCTTCACCTTGGGGCGCGCCCT 12430
 272 ACATGACGCCCAAGATCTCGGAGGCGTACACAGGTGTACAGGGCAATTGGCGCATCA 331
 12429 GCGCGCGGACACAGCTCGGAAAGGTGCGCGGACGACAGCGCTGCGGCTGACTTGCAGC 12370
 332 GCGTGGGGCGGTGGCGGACGACCCGCGCGCGCGCGCGCGCTGAGACTGTGAGCAAGCGC 391
 12369 CCGGCGACGCGCTGCTGCTGCGGTGCGGTGACGCGCGCGCGCGCGCGGCTGCGGAGCGC 12310
 392 GCTCTTACCGTTTGAAGCGGTGCGCGGACTACCGGAGGTGCGGGGCGGCGGCGGCGG 451
 12309 AGGTCTTACAGTTTCA---GCGCACCAAGGCGCTTCAAGCGCTTGAAGGGGCGCGCTCG 12253
 452 CCGTCAACGCGCGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 511
 12252 TCACCGGACGCGCGGACCTCGCGCGCGGATCGCGCGCGCTTCACACTTGGCTTGCAGC 12193
 512 TGGAGGCA 571
 12192 TGGCGGCGGCGGCA 12133
 572 TGGCGGCGGCA 631
 12132 TGGGCGCTCAACCTCCCTCGAGCGGTTCCGAGGTCAATCGAGCGGAACCGCGCGACACG 12073
 632 CCACGAGAGATCATCGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 691
 12072 CCGCGCTTACCGGACGCTCGGCGGACCTCCCGGGCGCTCTCGTCCGCGCGCGCGCGCG 12013
 692 CCACATGCGCGGACGAGCGGACGACATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
 12012 AGGCGCTCAACACCAACGATGATGATGATGATGATGATGATGATGATGATGATGATG 11953
 752 GCGGAGAGTGTGAGCTGCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 811
 11952 ACCGGGACCTGCTCATGAGGTCTCGAAGCGCGAAGGGGTGACACCGCGCGCTTACTTCT 11893
 812 TGGCGCGCTGCGGAGAGCGTGCCTGATGAG---GCGCGCATGGGCGCGCGCGCGCGCG 862
 11892 CCGCGCGGTGCGGAGAGTGTGAGCGGTACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCA 11833
 863 AGGAGGTCTGCGTGTGAGCGGTATCTCTACGTGCGCGGTCCACACATGCTCAAGAGAGCG 922
 11832 CCGAAGCGCTGCGCGCGCGGTGCTGTCTCTCGCACCGGACCGCGCATGCGGCGAGCGAG 11773
 923 GCGTCTCTGCGCGGACGACACATCATGATGATGATGATGATGATGATGATGATGATGATG 982

DB 11772 ACATCCGCGGCTCGCGGACGACCTGCTGCTGTGCGGACCGCGCGCGGACGATGACG 11713
 983 GCGCGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1015
 DB 11712 CCGCGGACCGGACGACG 11680.

RESULT 7
 US-09-861-289-3/c
 Sequence 3, Application US/09861289
 Patent No. US20020110897A1
 GENERAL INFORMATION:
 APPLICANT: Sherman, D.H.
 APPLICANT: Liu, H.
 APPLICANT: Xue, Y.
 APPLICANT: Zhao, L.
 TITLE OF INVENTION: DNA encoding methymycin and plitromycin
 FILE REFERENCE: 600,438US1
 CURRENT APPLICATION NUMBER: US/09/861,289
 PRIOR FILING DATE: 2001-05-18
 PRIOR APPLICATION NUMBER: 09/105,537
 NUMBER OF SEQ ID NOS: 43
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 3
 LENGTH: 13613
 TYPE: DNA
 ORGANISM: Streptomyces venezuelae
 US-09-861-289-3

Query Match 4.7%; Score 81.8; DB 10; Length 13613;
 Best Local Similarity 44.2%; Pred. No. 1.1e-09;
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

32 CCATGTCCTCCATCCGATTCCTCCAGCCCTTCGACCTGCTCTGCGGGCGGCAACCTCATCG 91
 12669 CCACCTGCAAGCGCACGGCGGCTCCAGCTCTCTGCGGACAGCCCGGCTTCACGGCGC 12610
 92 ACGGCAAGCAACACCCCGGGGGGCGCGCGGACCTGGGGTGGCGGCGGACCGCATCGCGC 151
 12609 AAGTATCATGCGCGTGCATGACGTTCGCGCCGACCGCCGACGACACTGCGTGGATCGCGC 12550
 152 CCATCGGGGATCTGTGCGAGCCCGCGCGGACACCCCGGGTGCAGTGTGCGGCGTGTGG 211
 12549 TCACCCGCGGTTCGCGGACATCGACCCGCGGACACCGGCAACCTCGAACCCGAGCAGGTGG 12490
 212 TCGGCGCGGCTTCATGACATCGGACACCCGACGACGACACTACTGCTCAGCGGTGCGG 271
 12489 CCGCGGGGTACACACCCCGGACCTCGGCGCTGCTGCGGCTTCACCTTGGGGCGCGCCCT 12430
 272 ACATGACGCCCAAGATCTCGGAGGCGTACACAGGTGTGACAGGGCAATTGGCGCATCA 331
 12429 GCGCGCGGACACAGCTGCGGAAAGGTGCGCGGACGACGAGCGCTGCGGCTGATCGAGC 12370
 332 GCGTGGCGCGGTGCGGACGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
 12369 CCGCGCGCGCTTCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 12310
 392 GCTCTTACCGTTTGAAGCGGTTCGCGGACTACCTGGAAGCGGTGCGGGCGGCGGCGCGG 451
 12309 AGGTCTTACAGTTTCA---GCGCACCAAGGCGCTTCAAGCGCTTGAAGGGGCGCGCTCG 12253
 452 CCGTCAACGCGCGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 511
 12252 TCACCGGACGCGCGGACCTCGCGCGGATCGCGCGCGCTTCACACTTGGCTTGCAGC 12193
 512 TGGAGGCA 571
 12192 TGGCGGCGGCGGCA 12133
 572 TGGCGGCGGCGGCA 631
 12132 TGGGCGCTCAACCTCCCTCGAGCGGTTCCGAGGTGATGATGATGATGATGATGATGATGATG 12073

```
OY 632 CCACGAGAGATCATCGAGGTGTGCGGCGGCTGAGCCGCGATGCGGCATCTACGCCA 691
    |||
Db 12072 CCGCTTACCGGACGACCTGCGGAGACTCCGCGCTCTCTGCGCCACGACCGCGCC 12013
OY 692 CCCACATGCGCGCAGAGCGAGCAGCATCTGCGCGCGCGCTGCGGAGAACCTTCCGATG 751
    |||
Db 12012 ACGGCTCAACAGACACAGCATGAGTGTGATGATGATGATGATGATGATGATGATGATG 11953
OY 752 GCGCGAGCTGAGCTGCGCGGTGTGTGATCTGCGACCAAGGTCATGAGGCCAGCCCAAT 811
    |||
Db 11952 ACCGCGACCTGCTGATGAGGTCTGTAAGGCGGAGCGGATGACACCGCGCTACTTCT 11893
OY 812 TCGCGCGCTGCGGAGAGCGCTGCGCGCTGATGAG-----GCCGCAATGCGCGCGC 862
    |||
Db 11892 CCGCGGCTGCGGAGAGCTGAGGAGCGGTACCGGCGGAGCGGAGCGCGCGCTGCGCGC 11833
OY 863 AGGAGCTGCTGCGGAGCGCTGATCCCTACGTGCGCGGCTGCGACCAATGCTCAAGCAGAG 922
    |||
Db 11832 CCGAGAGCTGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCGACCGCGCGCGCGCGAGC 11773
OY 923 GCGTCTCTGCGCGGAGCGACGACCATCATCTGCTGCGAGCGCTTCCCGAAGTGAAGC 982
    |||
Db 11772 ACATCTGCGCGGAGCGCGCGCGCGCTGCTGCTGCTGCGCGAGCGCGCGCGCGAGC 11713
OY 983 GCGCGAGCTGAGTGAATCTGCGCGCGCGAGCGCG 1015
    |||
Db 11712 CCGCGACCGCGGACGCGCGCGCGCGCGCGCTG 11680
```

RESULT 8

```
US-10-166-087-45
; Sequence 45, Application US/10166087
; Publication No. US20030077767A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia Biosciences Inc.
; APPLICANT: Farnet, Chris
; APPLICANT: Stafie, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthracyclin
; FILE REFERENCE: 3014-205
; CURRENT APPLICATION NUMBER: US/10/166,087
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 45
; LENGTH: 4341
; TYPE: DNA
; ORGANISM: Streptomyces refinesus subspecies thermotolerans
US-10-166-087-45
```

```
Query Match 4.58; Score 79.2; DB 9; Length 4341;
Best Local Similarity 41.9%; Pred. No. 4.6e-09;
Matches 474; Conservative 0; Mismatches 658; Indels 0; Gaps 0;
```

```
OY 31 TCCATGTCCTCAATCCGATTCCTGACCTGCTGCTGCGGCGCGACCTCTATC 90
    |||
Db 595 TCGCTGCTCCACTGCTGAGACATGCTGCGGCGCGCGCGCGCGCGCTGCTG 654
OY 91 GACGCGACGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 150
    |||
Db 655 CCGGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 714
OY 151 GCGCGAGCTGAGAGCTGCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 210
    |||
Db 715 GCGCGAGCTGAGAGCTGCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 774
OY 211 GTCGCGCGCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 270
    |||
Db 775 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 834
OY 271 GACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 330
    |||
```

```
Db 835 GTCACCTGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 894
OY 331 AGCTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
    |||
Db 895 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954
OY 391 GCGTCTTACCGCTTTCGAGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 450
    |||
Db 955 GACATCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014
OY 451 GCGCTTACCGCGCGCGCTTTCGAGCGCTTTCGAGCGCGCGCGCGCGCGCGCGCG 510
    |||
Db 1015 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074
OY 511 TTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570
    |||
Db 1075 ATGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1134
OY 571 ATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630
    |||
Db 1135 GTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1194
OY 631 ACCACGAGAGATCATGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690
    |||
Db 1195 ACCTGCTGCTGAGAGCACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1254
OY 691 ACCACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750
    |||
Db 1255 GCGCGCGCGCTGAGAGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1314
OY 751 GCGCGGAGCTGAGAGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810
    |||
Db 1315 CCGCTGCGCGCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1374
OY 811 TTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 870
    |||
Db 1375 CAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1434
OY 871 TCGCTGAGCGCTATCCCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 930
    |||
Db 1435 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1494
OY 931 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 990
    |||
Db 1495 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1554
OY 991 CTGAGTGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1050
    |||
Db 1555 CCGGTGCGCGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1614
OY 1051 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1110
    |||
Db 1615 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1674
OY 1111 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1162
    |||
Db 1675 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1726
```

RESULT 9

```
US-10-166-087-1
; Sequence 1, Application US/10166087
; Publication No. US20030077767A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia Biosciences Inc.
; APPLICANT: Farnet, Chris
; APPLICANT: Stafie, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthracyclin
; FILE REFERENCE: 3014-205
; CURRENT APPLICATION NUMBER: US/10/166,087
; NUMBER OF SEQ ID NOS: 51
```

3349 CGTGTACGCCAAGAACCACTGGATCCGCGAGACGTCTGCGGGTCTGCTCGAGCGCGG 3290

QY 780 CTCGACACACAGGTGATGAGCCAGCCCAATTTCGGCCGCTCGCGGAGACGCTGCCCT 839
 Db 3289 CGGGAACCCAAAGCCCTCCGACATGCAACGGCTTACGCCGCTGCGGGCTGTGTGCT 3230
 QY 840 GATCGAGGCGCCATGAGCGCCAGACGCTCTCGTGGACGCGTATCCATCGTGGCGCG 899
 Db 3229 CGGCGGGGCTCCGCGCGCGTCTGAGACATGATGCTCGCGCGCGCGCGCGCTGCTG 3170
 QY 900 CTCACCATCTCAAGAGAGACCGCGCTGCTGCGCGAGACACATCATCCTGGTG 959
 Db 3169 CGTGAAGCGCTAGCGGGGCAACAGCTGACACACCTCGCGCGAGCGCAAGATCTCGA 3110
 QY 960 CAAGCCCTTCCCGACATGAGCGGGCGCGACCTGAGTGAAGTCCGCGCGAGCGCGCA 1019
 Db 3109 GGGGATGCTCGACACGCTCACGGGCTTAGCGGTGA-----CCCGCGCGCTAGAGCG 3056
 QY 1020 ATCCAAATACAGCTGAGTGGTCCGAGCTGACCGCGCGCGCATCTTCACTCATGATGA 1079
 Db 3055 GTGTGGGAACACCATGTGTGACTACATGAGGAGCTTACGGAGCTGCGAGCGCGCTGCT 2996
 QY 1080 CGAACCCGACGTGACAGCGCATCTGCGCTTCCGCGCGACCATGATGCTCCGAGCGCT 1139
 Db 2995 GGACTTGTGCTGAGCGCGCGCTGAGACCTGGAACCTGCGCAACAACCTGACAGCGC 2936
 QY 1140 GCCGACGACGACGCGCGCATCCGCGCTGAGGAGCATTCGCCGGGGGCTGCGGGCA 1139
 Db 2935 GCGGACCGCGCG---CGGCTTTCACGACCAAGCGCGCTGCTGCGCGCTGAGTGG 2879
 QY 1200 CATGCGCGGACCTTGGCGCTTTCGCCGAGAGCGCGGTATGAAATGACCGCGCT 1259
 Db 2878 CGGAGGCTGACACATGATGCGGCTGCGGCTATGCGGGTTTCGGAATGCTGCGCGC 2819
 QY 1260 GACCGCGCGCGCTTGGCGCTGCGCGCGCGCGCGGACGCTGAGCGCGGTACTTCCGCA 1319
 Db 2818 CAAACAAGTGGCGCGACGCGCGCGCTGCTGCGCGCGCGCGCGCGCGAGGCTGCTGC 2759
 QY 1320 CCTGTGTGTTC 1332
 Db 2758 GCGCGCGCTGATC 2746
 RESULT 11
 US-09-860-846-1
 : Sequence 1, Application US/09860846
 : Patent No. US2002016742A1
 : GENERAL INFORMATION:
 : APPLICANT: Sherman, D. H.
 : APPLICANT: Liu, H.
 : APPLICANT: Xue, Y.
 : APPLICANT: Zhao, L.
 : TITLE OF INVENTION: DNA encoding methymycin and plikromycin
 : FILE REFERENCE: 600,438US1
 : CURRENT APPLICATION NUMBER: US/09/860,846
 : PRIOR FILING DATE: 2001-05-18
 : PRIOR APPLICATION NUMBER: 09/105,537
 : NUMBER OF SEQ ID NOS: 43
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 1
 : LENGTH: 15872
 : TYPE: DNA
 : ORGANISM: Streptomyces venezuelae
 US-09-860-846-1
 Query Match 4.2%; Score 73.8; DB 9; Length 15872;
 Best Local Similarity 43.2%; Pred. No. 6.3e-08;
 Matches 586; Conservative 0; Mismatches 747; Indels 24; Gaps 4;

Db 4859 GCCCTGACCGCTACCTGACCGCGCGCTGGGAGATGCTGGCGCGGAGAGACTG 4918
 QY 295 GGGTCACACAGGCTGATGACGCGGCAATTGCGGATACACTTGGCGCGCTGCGCACGC 354
 Db 4919 GACCTCAACCGGTACACACACCGCGCTTCTGCGCTTTCGAGTGGCGCTTCCGCTC 4978
 QY 355 AACCGCGCGCGCGCTGACCTGATGAGAGCGCGCTTACCGTTGACCGCTTC 414
 Db 4979 CTGACACACACGCGCTGCTCCGACCTGCTCACCGGACACTCTCGTGGGAGATGCGC 5038
 QY 415 GCGGACTACCTGAGCGGCTTCCGAGCCACGCGCGCGCTCAACCGCGCTGATGATG 474
 Db 5039 GCGCGGACGTCGCGCGGCTCTCTCCCTGACGACGCGGAGCTCTGTCACCGCGCGC 5098
 QY 475 GGGCATTCACAGCTGCGCGCGCGCTGATGCGGACTTTCAGCGCGCGCGCACGAG 534
 Db 5099 GGGCGGCTCATGACGTGCGCGCGCGGAGCGCGCGATGATCGCTTGCAGCGCGCGAG 5158
 QY 535 GAAATCGCGGCTATGCGGAGACCTGCGGAGAGACATGCGGACGCGCGCATGCGGAT 594
 Db 5159 GCGAGGTGCTGAGTCCCTGAGAGGCTACGAGGCGAGGTCGCTGCGCGCTGAC 5218
 QY 595 TCGACGCGCGCTTCTACCGCGCGCGCGCGCGCGCACACGAGAGATCATGAGTG 654
 Db 5219 GGAACCGCGCGCTGCTGCTCCGCGAGCGGAGCGCGCGGAGATCCGCGCGCTA 5278
 QY 655 TCGCGCGCGTGAAGCGGCGGATGCGGCGATCTACGCGCACATGCGCGAGAGAGAG 714
 Db 5279 TGGCGGAGACGCGCGCGCGCACCGCGCGCTGCGCTGACCGACCGCTTCACTCCG 5338
 QY 715 CACATCTGCGCGCTGAGAGAACTTCCGATCGCGCGCGGAGTGGAGTGGCGGT 774
 Db 5339 CACATGACGACGTCCTGAGAGGTTCTCCGCGTCCGAGGCGCTGACCTTGCAGAG 5398
 QY 775 GTGATTCGACACCAAGTATGAGCGGACCGCAATTGCGCGCTGCGCGAGAGCTG 834
 Db 5399 CCGGAGATCCCGCTGCTGCTGACGCTGACCGCGCGCTGCTGACGCTCGCGCA----- 5451
 QY 835 CCGGTATGAGGCGCGCATGCGCGCGCGCGGACGCTGCTGAGACCGTATCCCTAGTG 894
 Db 5452 --GCTACCTGCGCGCGCTGCTGAGTGAACCAATCGCGCGCTTCTGAG 5509
 QY 895 GCGGCTCACATGCTCAAGACGAGACCGGCTGCTGCGCGGAGACACATCATCAC 954
 Db 5510 GCGGTCCGACCTTGGCGCGCGGAGAGCGACCGCTCTGCTGAGATCGCGCGCGCG 5569
 QY 955 TGTGCAAGCTTCCCGCAATGAGCGGCGGAGCTGATGATGCGCGCGAGAG 1014
 Db 5570 GTCTCACGCGACTCGCGGAGAGCTTTCGCGCGGACGAGAGCGCGCGCGCG 5629
 QY 1015 GGCAAATCCAAATGACAGTGGTCCCGAGCTGACCGCGCGCGCGCATCTACTCATG 1074
 Db 5630 GACGTACAGGTCTCCCGCTGCTGCGCGCGCGCGCGCGCGAGCCCTCGCGCG 5689
 QY 1075 ATGAGCAACCCGACGATGAGCGCATCTGCGCTTGGCGCGGACATGATGCGCTCGAG 1134
 Db 5690 GGTCTGCGACCGCGCATGTCACGCGGACACCTTGGACCGCGCGCTTCTCCGAG 5749
 QY 1135 GCGCTCCGAGAGAGCGCGCGCATTCGCGCTTGGGAGCACTTCCGCGGCTG 1194
 Db 5750 GGGCGCGGACGAGACTCGCGCATGCTTCCGCGGAGACTATGCGGTGAGCGCC 5809
 QY 1195 GGGCAC-----TAGCGGCGGACCTGCGCGCTTCCGCTGAGAGCGCGGATGAGAG 1248
 Db 5810 GAGGCTCGTACGAGCGCGCGGACTCGGCTTTCGACCGCGCGCGCGCTGCTGAG 5869
 QY 1249 ATGACGCGCTTACCGCGCGCGCTTCCG-----CTGCGCGGCGGCGGAGCTGAG 1302
 Db 5870 ACCAGGCTGAGTTCGCGCGCGGAGCGCGCTGCTGACCGCGCTTCTCCCTGAGC 5929
 QY 1303 GCGGCTACT---TCGCGACCTGCTGCTGCTGACCGCGCGCGGAGGCTGATACCGC 1359

Db 5930 GACCAACCCCTGCTGGCCGACACATGCTCAACGGCCGCTCTGTTGCGGCCACCGCC 5989
 QY 1360 ACCTTGACAAACCCCTACAGAGCGCGCCGCGCATTCCTGTTAGCTAAGGGCGG 1419
 Db 5990 TTCTTGAGAGCTGCGCTCTGCGCGGCGGACACCTGCGGGGCTCCGGGTGAGGAACTC 6049
 QY 1420 CCGGCTTGCGAAGAGAGAGGCTTACCGCGGACAGCATGGCGCGGCTGCTGACGACG 1479
 Db 6050 ACCCTCGAAGGCGCGCTGCTGCTCCGACGAGGGGCGCGCTCCGATCAGAGTGGCGTG 6109
 QY 1480 GCGGCTGAGCGCGCGCGGCGGACCTTACATCCGCGG 1516
 Db 6110 AGCGGCGAGCGGAGTGGCGCGCGGCGGCGACCTTCG 6146

RESULT 12

US-09-988-384B-1
 ; Sequence 1, Application US/09988384B
 ; Publication No. US20030073824A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and plikromycin
 ; FILE REFERENCE: 600.536US1
 ; CURRENT APPLICATION NUMBER: US/09/988,384B
 ; PRIOR FILING DATE: 2001-11-19
 ; PRIOR APPLICATION NUMBER: PCT/US99/14398
 ; PRIOR FILING DATE: 1999-06-25
 ; PRIOR APPLICATION NUMBER: US 09/105,537
 ; PRIOR FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 53
 ; SEQ ID NO 1
 ; LENGTH: 15872
 ; TYPE: DNA
 ; ORGANISM: Streptomyces venezuelae
 ; US-09-988-384B-1

Query Match

Best Local Similarity 4.28; Score 73.8; DB 9; Length 15872;
 Matches 586; Conservative 0; Mismatches 747; Indels 24; Gaps 4;

QY 175 GCGCGCACACCCGCTGCACTGTGCGGCTGTGTCGCGCGCGCTTCATGACTCG 234
 Db 4799 GCGGAGGAATGTCGCGCGCGCGGACCCCGTTCGCGCGCGCTTCGACACCGTGTAGCG 4858
 QY 235 CACACCCAGAGAGACAATACTGCTGAGCGCTGCGACATGAGCGCCAAGATCTCGAG 294
 Db 4859 GCGCTGACCGCTGACCTCGACCGCGCGCTGCGGAGATGCTGCGCGCGCGGAGAGACTG 4918
 QY 295 GCGGTACCAAGGTGTGTCAGGGGCAATTGCGGATAGCGCTGCGCGCTGCGGACGCG 354
 Db 4919 GACCTACCGCGGTACACCGACCGCGCTTCCTGCGCTTCGAGGTGCGGCTGCTCCGCTC 4978
 QY 355 AACCCGCGCGCGCGCTGAGCTGTGAGAGAGGCGGCTTACCGGTTTCAGAGCGCTTC 414
 Db 4979 CTGGAACACAGAGCGCTGCTCCCGACGCTGCTACCGGCACTCCGTGCGCGGATGCGC 5038
 QY 415 GCGGACTACCTGAGCGGCTTGGCGGCGACCGCGCGGCTCAACCGCGCTGTATGCTG 474
 Db 5039 GCGGCGCACGTCGCGGTGCTCTCTCTGAGAGAGCGGCAAGTCTGTCACCGCGCGC 5098
 QY 475 GCGCATTAACGCTGGCGCGCGCGGCTGATGCGGACTTCAGAGCGCGCGCGGACAGAG 534
 Db 5099 GCGCGGCTCATGAGTGGCGCGCGGCGGCGGCGCGGATGATGCGCTGACGCGGCGAG 5158
 QY 535 GAATGCGCGGCTGCGGAGCTGCGCGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCT 594
 Db 5159 GCGGAGGTGTGAGTCCCTGGAAGGCTGAGAGGCGAGGAGGTGCGCGCGCGGCTCAAC 5218
 QY 595 TGACCGGCGGCTTACCGCGCGCGCGCGCGGCGGCGGCGGCGGAGAGATTCATCGAGGTG 654

Db 5219 GACCCACACCGCGCTGCTGCTTCGCGGACGCGGACGCGCGCGGAGAGATCCGCGCTA 5278
 QY 655 TGCGCGCGCTGAGCGCGCATAGCGGCACTTACGCGACCCACATGCGGAGAGAGGAG 714
 Db 5279 TGCGCGGAGAGCGCGCGCGGACCGCGAGGCTGCGCGCTGACACAGCGCTTCACCTCCCG 5338
 QY 715 CACATCGTGCGCGGCTGAGAGAACTTCGCGATCGCGCGGACGCTGAGTGGCGGTG 774
 Db 5339 CACATGAGAGAGCTGCTGAGAGATTCCTCCGGGCGCGGAGGCGCTACCTTGAAGAG 5398
 QY 775 GTGATTCGACACAGAGTATGAGCGGACCGCAATTTGCGCGCTGCGCGGAGAGCTG 834
 Db 5399 CCGGAGATCCCGCTGCTTCACGCTACCGCGCGCGCTGCTGACGTCCGCGGGA----- 5451
 QY 835 CCGCTGATCGAGCGCGCATGCGCGGACGCGGACGCTGCTGAGAGCGGTATCCCTAGTG 894
 Db 5452 --GCTCACTGCGCGCGGCTGAGTGGGTGACACAGATCCGCGCGCGGCTTCCTGAGAC 5509
 QY 895 GCGGCTTCACCATGCTCAAGCAGAGACCGCGCTGCTGCGCGGACGCGACCATCATCAGC 954
 Db 5510 GCGCTCCGACCGCTGCGCGCGGCGGACGCGACCGCTCTGCTGAGATCGGCGCGGAGCC 5569
 QY 955 TGTGCAAGCGCTTCCCGAAGTACGAGCGGCGGCGGACCTGATGAATCGCGCGGAGCGC 1014
 Db 5570 GTCTTCAGGCACTGCGCGGAGAGGCTGCGCGCGCGGACGAGCGCGCGGAGCGCGCG 5629
 QY 1015 GCGAATATCAAGTACGAGAGTGTGCGCGAGCTGAGCGCGCGGCGGCGGCGCATCTTATG 1074
 Db 5630 GAGGTACAGGTGCTGCGCGGCTGCTGCGCGGCGCGCGCGCGCGGCGGCGGCGCGCG 5689
 QY 1075 ATGAGCAGACCGCGAGCTGACAGCGCATCTGCGGCTGCGCGCGGCGGCGGCGGCGG 1134
 Db 5690 GGTCTTCGAGACCGCGCGCTGCTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGAC 5749
 QY 1135 GCGCTGCGGACGAGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1194
 Db 5750 GCGGCGCGGACGAGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5809
 QY 1195 GCGGCGC-----TATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1248
 Db 5810 GAGCGCGGCGGAGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5869
 QY 1249 ATACCGGCTGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1302
 Db 5870 ACCACGCTGAGGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5929
 QY 1303 GCGGCGGACT---TGGCGGCGGCTGTGCTGTGAGACCGGCGGCGGCGGCGGCGGCGG 1359
 Db 5930 GACCAACCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5989
 QY 1360 ACCTTGAAACACCTTACCGGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1419
 Db 5990 TTCTTGAGAGTGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6049
 QY 1420 CCGGCTTGCAAGAGAGAGCGGCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1479
 Db 6050 ACCCTCGAAGCGCGGCTGCTGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 6109
 QY 1480 GCGGCGCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1516
 Db 6110 AGCGGCGAGCGGAGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6146

RESULT 13

US-09-861-289-1

; Sequence 1, Application US/09861289
 ; Patent No. US20020110897A1

GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and plikromycin

Db 1668 GGGGGCCGGCCGCTTCGGCTGGGGGCTGGGGCCAGCGCGCGCCGCTTGGCCATGAGCCG 11727
 QY 369 CTTGAGACTTGTGAGCAGAGCGCGCTTACCGCTTTCAGAGGCTTGGCCATACCTGGA 428
 Db 1728 CCGATACGACCGCGCAGAGAGGCTTCTGCTGACCACTGGCGCGCCCTACAGCGGC 1787
 QY 429 CGGTTGGGGCCACGCGCGCGCTCAAGCGCCCTGTATGATGGGCCATCAACGCT 488
 Db 1788 CTTGTTGGCGCGGAGAACCGCGCTGACGGGGGCGCGGGAGCCCGCGCCGCGCGC 1847
 QY 489 GCGCGCGCGCTCATGCGCGGACTTGCAGCGCGCGCCGACGAGAGAAATCGCGCCAT 548
 Db 1848 AGATGAGAGAGGGGCTGGCGCCCTCCCGCGCGCCGACGCGGCGAGCGCGCTGCCCG 1907
 QY 549 GCGGAGCTTGGCCGAGAGAACCTTGCAGCGCGCGCTTGCAGCTTTCAGCGCGCTT 608
 Db 1908 CCGGTAAGCGCGCGCGGAGTCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCTC 1967
 QY 609 CTACCG 662
 Db 1968 CCG 2027
 QY 663 GCTGAGCGCGCATGCGCGCATGTACGCGCACCAATGCGCGAGAGGAGGAGCACTGCT 722
 Db 2028 GCG 2087
 QY 723 GCGCGCGCTGAGAGAACTTCCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 782
 Db 2088 GAGAGCGCTGGCGCGAGGCTTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2147
 QY 783 GCACCAACAAGTATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842
 Db 2148 CCG 2207
 QY 843 CGAGCGCGCGCATGCG 902
 Db 2208 GCACCG 2267
 QY 903 CACACTGCTCAAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 962
 Db 2268 GCTGCTGCTCATG-----CGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2321
 QY 963 GCGCTTCCCGCACTGAGCG 1022
 Db 2322 CGCGCTGGCG 2381
 QY 1023 CAATGACGACTGTGCG 1082
 Db 2382 GCGGGAGCG 2441
 QY 1083 ACCGAGCTGACAGCG 1142
 Db 2442 CCAAGACTGTGCG 2501
 QY 1143 GCAAGCAGAGCG 1202
 Db 2502 GCG 2561
 QY 1203 TGGCGCGAGCTTGGCG 1262
 Db 2562 GGGCG 2621
 QY 1263 GCGCGCGCGCTTGGCG 1322
 Db 2622 CGCGCGCGCGCTGGAGCG 2681
 QY 1323 GGTGTGTTGACCG 1376
 Db 2682 GGGGCGCGAGCG 2741
 QY 1377 CGAGCG 1436

Db 2742 CCG 2801
 QY 1437 GCGCTTACCG 1496
 Db 2802 GCGCGCGCGAGCG 2861
 QY 1497 CCAAGCTTACAACTCCG 1541
 Db 2862 CTACTGCTCTCCCG 2906

RESULT 15
 US-09-815-242-4056
 Sequence 4056, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlson, Karl L.
 APPLICANT: Zykkind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELTRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 4056
 LENGTH: 3189
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-815-242-4056

Query Match 4.1%; Score 72.4; DB: 10; Length 3189;
 Best Local Similarity 43.8%; Pred. No. 1.5e-07;
 Matches 371; Conservative 0; Mismatches 471; Indels 6; Gaps 1;

QY 16 GCGGAGAGAGATTTCATGTCATTCCTCAATCCAGCTTCCAGCTTGCATGCTGCTGCGG 75
 Db 319 GTGACAGTGCAGAACCGCTCAGTCCCGAGCGAGCCGAGAGCTCCCGAGAGAGTACCCG 378
 QY 76 GCGGCGACCTTCATGACGCGAGCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 135
 Db 379 CTGCGGATACCGCTGACAAAGGCTTGCAGCACTTACCATGATGCTGACCTGACCTCG 438
 QY 136 GCGGACCGCATGCGCGCATGCGCGATGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 195
 Db 439 CCGGATACCGCTGACAGACATGCTACCTGTGAGACTAGCGCGCTGCTCACTGAGAGAC 498
 QY 196 GGTGTGCGCTGTGCTGCGCGCGCGCTTACATGACTGACACACCGACGACGACAATAC 255
 Db 499 GAACTGCGCGCGCTGAGAGCGCGCTGCGCGAGCTGCAATGTTGCGCTGCGCGCAATATCG 558
 QY 256 CTGCTCAGGCGTGGGACATGACGCGCAAGATCTTGGCGAGGCGTGTACCAACGCTGTCACG 315

